UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DEPARTAMENTO DE MICROBIOLOGIA E IMUNOLOGIA

Molecular and Genomic Epidemiology of Multidrug Resistant *Mycobacterium tuberculosis* Isolates in Lisbon, Portugal

João Ruben Lucas Mota Perdigão

DOUTORAMENTO EM FARMÁCIA

MICROBIOLOGIA

2013

UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

DEPARTAMENTO DE MICROBIOLOGIA E IMUNOLOGIA

Molecular and Genomic Epidemiology of Multidrug Resistant *Mycobacterium tuberculosis* Isolates in Lisbon, Portugal

João Ruben Lucas Mota Perdigão

Tese orientada pela Prof. Doutora Isabel Portugal

Tese especialmente elaborada para a obtenção do grau de doutor em Farmácia, especialidade de Microbiologia.

2013

Todas as afirmações efetuadas no presente documento são da exclusiva responsabilidade do seu autor, não cabendo qualquer responsabilidade à Faculdade de Farmácia da Universidade de Lisboa pelos conteúdos nele apresentados.

João Perdigão teve o apoio financeiro da Fundação para a Ciência e Tecnologia através de uma bolsa de Doutoramento (SFRH/BD/45388/2008).

"If it is a terrifying thought that life is at the mercy of the multiplication of these minute bodies, it is a consoling hope that Science will not always remain powerless before such enemies..."

Louis Pasteur

ACKNOWLEDGEMENTS

I would like to acknowledge all those that in some way have contributed to the work that is here presented:

First of all, to Prof. Doutora Isabel Portugal that back in 2005 has received me in her laboratory, where I have remained since. Thank you very much for your friendship, constant availability, guidance and, to willingly undertake new projects and ideas!

To Professor Doutor José Moniz-Pereira, for allowing me to use the URIA facilities to conduct the great majority of the work that is presented here and, for all the scientific discussions and advices.

To Prof. Doutora Laura Brum for receiving me in her team, in the somewhat distant year of 2006, and allowing me to conduct the experimental work for some these studies in the Mycobacteria Unit of the National Institute of Health Dr. Ricardo Jorge. Thank you very much.

To Prof. Doutor Miguel Viveiros and Prof. Doutora Isabel Couto for providing clinical strains and data. Furthermore, my sincere gratitude for all the incentive, support and stimulating scientific discussions.

To Doutora Luísa Jordão for providing clinical strains and for the assistance with MIRU-VNTR typing calibration.

To Prof. Doutor Taane Clark for opening for me the doors to the whole-genome sequencing world! Thank you very much for all the support and for sharing with us the more complex bioinformatic scripts.

To Rita Macedo, for her inestimable help in my first years, for the early MIRU-VNTR genotyping data and, more recently, for providing clinical isolates and phenotypic resistance data. Thank you very much!

To Carla Silva, with whom I first started to uncover the phylogenetic positioning of Lisboa strains. Thank you very much for your help and friendship.

To Claudia Pinto, for sharing with me our early venture in insertion site mapping. Thank you very much for your friendship and for your constant availability in creating and maintaining all the necessary working conditions in our unit.

To Hugo Silva for his recent but, nonetheless, great help in the laboratory conducting MIRU-VNTR genotyping, and therefore, contributing to the data presented here.

To Inês João with whom I had the pleasure of sharing molecular drug resistance studies some years ago, and for contributing with molecular data on isoniazid resistance.

To all the staff that a few years ago were working in the Mycobacteria unit at INSA, for their help, whenever it was necessary, and support: Dra. Elizabete Fernandes, Paula Cristovão, Lúcia Reis. Also, to the memory of Dra. Ana Garcia.

To all the students that have passed through the mycobacteria laboratory for their help in the laboratory work: Ana Rute, Ana Catarina, Catarina Milho, Ana Isabel, Ricardo, Cátia, Eduarda, Vânia and Diogo.

To all with whom I have had the pleasure of working alongside, sharing the MM140 laboratory: Helena Barroso, Inês Bartolo, Pedro Borrego, Cheila Rocha, Andreia Martins, Rita Calado, Joana Duarte, Francisco Martin, Ana Rita Diniz, Claudia Palladino, Hugo, Daniela and Marcelino.

I would also like to acknowledge all the professors in the Microbiology department for the incentive, availability and foremost, for maintaining a very human and friendly working environment: Prof. Doutor José Miguel Pereira, Prof. Doutor João Gonçalves, Prof. Doutora Aida Duarte, Prof. Doutora Madalena Pimentel, Prof. Doutora Graciete Freitas, Prof. Doutora Elsa Anes, Prof. Doutora Maria Manuel Lopes, Professora Doutora Odette Ferreira, Prof. Doutor Nuno Taveira, Prof. Doutor Carlos São-José, Dra. Quirina Santos Costa, Dr. João Vital,

Dra. Paula Resende. Special thanks to Prof. Doutora Madalena Pimentel for the help with my early work with mycobacteria cloning and plasmidic site-directed mutagenesis.

To Diana Machado, for her precious collaboration in exchanging data and mycobacteria semi-quantitative drug susceptibility testing.

To all other colleagues working in CPM-URIA: Marta Calado, Marta Gíria, Mariana, Paula Brito, Nuno, David, Soraia, Luís, Catarina, Ana Catarina, Lídia, Andreia, Daniela, Paula, Camila, Maria, Cátia, Filipa and Maria João.

A special acknowledgement to Lena Brás, Lídia and Teresa for all the help and availability to perform what apparently is simple laboratory work but simultaneously essential so that everything else can be done. Thank you very much!

To Ofélia and Vera for taking care of all the paperwork and for their constant good mood.

All the work presented herein wouldn't also be possible without the work of those that maintain the all the laboratories and the P3 facility in normal working conditions. D. Fátima, D. Noémia, Cleida and Sr. Augusto, thank you very much!

Um agradecimento especial aos meus sogros e avós por todo o apoio, incentivo e, por estarem sempre disponíveis!

Aos meus pais e irmão pelo apoio, incentivo e por terem criado todas as condições que me permitiram hoje chegar até aqui. Obrigado!

À Ângela, que estará sempre comigo! Este espaço é curto para expressar toda a minha gratidão por todo o apoio e compreensão demonstrado ao longo destes últimos anos e especialmente nesta reta final!

Dedico esta tese à memória da minha Avó Aida.

PREFACE

The work that is detailed in this thesis is centered on the theme of multidrug resistant tuberculosis in Portugal. We have aimed at obtaining a better picture of the epidemiology and drug resistance in Lisbon Health Region through the use of molecular biology techniques. This research was mainly conducted in the Unidade dos Retrovírus e Infecções Associadas from the Centro de Patogénese Molecular in Faculdade de Farmácia da Universidade de Lisboa, although a part of it was carried out in the Instituto Nacional de Saúde Dr. Ricardo Jorge.

This research theme is presented along this thesis as a compilation of scientific articles (Chapters 2-7) published in peer-reviewed international journals or yet in a stage of submission/preparation. These scientific articles and its order reflect the evolution in our knowledge concerning strain clonality and the molecular determinants of resistance in Lisbon Health Region.

A General Introduction (Chapter 1) precedes these chapters and is intended to present a global view on some of the tuberculosis topics and its current state of knowledge, which the author found necessary in order to provide the reader with a broader understanding of the research that is discussed in the subsequent chapters. A General Discussion (Chapter 8) follows all chapters as it is intended to summarize the main findings in the previous chapters and to give these some perspective in the light of the evolving knowledge.

This thesis is therefore based in the following manuscripts already published or in preparation:

Perdigao J, Macedo R, Joao I, Fernandes E, Brum L, Portugal I. Multidrug-resistant tuberculosis in Lisbon, Portugal: a molecular epidemiological perspective. *Microb Drug Resist* 2008; **14**: 133-43.

Perdigao J, Macedo R, Malaquias A, Ferreira A, Brum L, Portugal I. Genetic analysis of extensively drug-resistant Mycobacterium tuberculosis strains in Lisbon, Portugal. *J Antimicrob Chemother* 2010; **65**: 224-7.

Perdigao J, Macedo R, Silva C, Pinto C, Furtado C, Brum L, Portugal I. Tuberculosis drug-resistance in Lisbon, Portugal: a 6-year overview. *Clin Microbiol Infect* 2011; **17**: 1397-402.

Perdigao J, Macedo R, Silva C, Machado D, Couto I, Viveiros M, Jordao L, Portugal I. From multidrug-resistant to extensively drug-resistant tuberculosis in Lisbon, Portugal: the stepwise mode of resistance acquisition. *J Antimicrob Chemother* 2013; **68**: 27-33.

Perdigao J, Macedo R, Machado D, Silva C, Jordao L, Couto I, Viveiros M, Portugal I. GidB mutation as a phylogenetic marker for Q1 cluster *Mycobacterium tuberculosis* isolates and intermediate-level streptomycin resistance determinant in Lisbon, Portugal. *Manuscript in preparation*

Perdigao J, Silva H, Machado D, Macedo, Maltez F, Silva C, Jordao L, Couto I, Mallard K, Coll F, Hill-Cawthorne G, McNerney R, Pain A, Clark TG, Viveiros M, Portugal I. Unraveling *Mycobacterium tuberculosis* genomic diversity and evolution in Lisbon, Portugal, a highly drug resistant setting. *Manuscript in preparation*

Other publications:

Published manuscripts:

Perdigao J, Macedo R, Ribeiro A, Brum L, Portugal I. Genetic characterization of the ethambutol resistance-determining region in *Mycobacterium tuberculosis*: prevalence and significance of embB306 mutations. *Int J Antimicrob Agents* 2009; **33**: 334-8.

Perdigao J, Milho C, Carrilho L, Brum L, Portugal I. Genotypic analysis of *Mycobacterium tuberculosis* isolates from a Lisbon hospital in Portugal. *Rev Port Pneumol* 2009; **15**: 761-9.

Veigas B, Machado D, **Perdigao J**, Portugal I, Couto I, Viveiros M, Baptista PV. Au-nanoprobes for detection of SNPs associated with antibiotic resistance in Mycobacterium tuberculosis. Nanotechnology 2010; **21**:415101.

Machado D, Couto I, **Perdigao J**, Rodrigues L, Portugal I, Baptista P, Veigas B, Amaral L, Viveiros M. Contribution of efflux to the emergence of isoniazid and multidrug resistance in *Mycobacterium tuberculosis*. *PLoS ONE* 2012; **7**: e34538.

Machado D, **Perdigão J**, Ramos J, Couto I, Portugal I, Ritter C, Bottger EC, Viveiros M. High-level resistance to isoniazid and ethionamide in multidrug-resistant *Mycobacterium tuberculosis* of the Lisboa family is associated with inhA double mutations. *J Antimicrob Chemother* 2013: *in press*.

Oral Communications in International Conferences:

Perdigão J, Malaquias C, Macedo R, Brum L, Portugal I. 2008. *Mutational characterization of tlyA gene in capreomycin-resistant Mycobacterium tuberculosis clinical isolates*. 29th Annual Congress of the European Society of Mycobacteriology. Plovdiv, Bulgaria

Perdigão J, Macedo R, Silva C, Pinto C, Furtado C, Brum L, Portugal I. 2011. *Six-year laboratory data report on drug-resistant tuberculosis in Lisbon Health Region, Portugal*. 21st ECCMID/ 27th ICC 2011. Millan, Italy.

Oral Communications in National Conferences:

Perdigão J, Sabino A, Milho C, Macedo R, Brum L, Portugal I. 2008. *Detecção de mutações no gene gidB em Mycobacterium tuberculosis resistentes à estreptomicina, na Região de Saúde de Lisboa*. XXIV Congresso de Pneumologia. Porto, Portugal

Perdigão J, Milho C, Furtado C, Carrilho L, Brum L, Portugal I. 2008. *Transmissão da tuberculose em Lisboa: Análise de uma população de estudo e caracterização por Mycobacterial Interspersed Repetitive Unit*. XXIV Congresso de Pneumologia. Porto, Portugal

Perdigão J, Macedo R, Silva C, Pinto C, Furtado C, Brum L, Portugal I. 2010. *Perspectiva Laboratorial da Tuberculose Resistente na Região de Saúde de Lisboa*. X & VIII Congresso Nacional de Doenças Infecciosas e Microbiologia Clínica & sobre SIDA. Coimbra, Portugal

RESUMO

A tuberculose (TB) multirresistente (MDR) suscita ainda sérias preocupações no âmbito de saúde pública. Apesar de Portugal registar uma tendência decrescente ao nível da incidência da TB nos últimos 20 anos, dados relativos à incidência de MDR-TB aparentam sofrer de alguma subnotificação quando comparados com os dados laboratoriais existentes.

Com o presente estudo pretendeu-se obter uma nova perspectiva do estado da TB em Portugal utilizando para tal técnicas de biologia molecular aplicadas ao estudo de estirpes circulantes de MDR-TB. Pretendeu-se focar este trabalho especificamente na Região de Saúde de Lisboa. Trata-se de uma região que face aos restantes padrões europeus apresenta uma incidência de TB considerada elevada, sendo neste aspecto apenas superada pela região do Porto. Estudos prévios realizados nesta região demonstraram a existência de um conjunto de estirpes geneticamente relacionadas, o qual foi designado por Família Lisboa.

São inicialmente descritas, de entre um conjunto de 58 isolados de MDR-TB, as mutações associadas à resistência a antibacilares de primeira linha mais prevalentes: C-15T na região promotora do operão *mabA-inhA*, para a isoniazida; S531L no gene *rpoB*, para a rifampicina; M360V no gene *embB*, para o etambutol; K43R no gene *rpsL*, para a estreptomicina; e, V125G no gene *pncA*, para a pirazinamida. Do ponto de vista genotípico verificou-se uma elevada prevalência do *cluster* Lisboa3, englobando 20.7% dos isolados analisados. Foi ainda registada uma elevada taxa de TB extensivamente resistente (XDR), correspondendo a cerca de 56% dos isolados MDR estudados.

Uma análise posterior de 26 isolados clínicos XDR, quer do ponto de vista genotípico quer do pondo de vista da base molecular da resistência, mostrou a existência de uma elevada clonalidade traduzida pela distribuição da totalidade dos isolados estudados por dois *clusters* genéticos distintos: Lisboa3 e Q1. Foram ainda identificadas mutações específicas, determinantes para a base molecular da resistência e associadas com os *clusters* descritos: foram identificadas mutações nos genes *tlyA* e *rrs* exclusivamente nos *clusters* Lisboa3 e Q1, respetivamente; as mutações D94G e D94A no gene *gyrA* foram identificadas apenas nos *clusters* Lisboa3 e Q1, respetivamente; e, uma mutação no codão 91 do mesmo gene foi identificada somente em isolados do *cluster* Lisboa3. Os resultados obtidos apoiam a hipótese de aquisição múltipla de resistência a antibacilares de segunda linha com expansão clonal na comunidade. A elevada prevalência de mutações no gene *tlyA* é ainda reveladora do importante papel destas como mecanismo de resistência à capreomicina na região tornando-se importante na tomada de decisão acerca da implementação de testes rápidos de diagnóstico molecular na região.

A análise dos registos de laboratório provenientes de um período de seis anos (2001-2006), aliada à genotipagem por MIRU-VNTR, permitiu verificar o nível elevado de XDR-TB na região, oscilando esta taxa entre valores próximos ao determinado inicialmente. Foi ainda observado um aumento gradual, de ano para ano, da fracção de isolados resistentes a todos os antibacilares de primeira linha. Verificou-se igualmente, entre 2004-2006, uma predominância das estirpes pertencentes aos *clusters* Lisboa3 e Q1 como principais responsáveis pelos casos de M/XDR-TB. Estes dados apontam para uma circulação contínua deste tipo de estirpes, evidenciando, juntamente com os dados moleculares de análise mutacional, um elevado nível de endemismo da M/XDR-TB na região.

Para tentar perceber melhor a dinâmica de aquisição de resistências ao nível dos antibacilares de segunda linha, bem como a base molecular da resistência a injectáveis de segunda linha foram ainda pesquisadas mutações no gene *eis*, *tlyA*, *rrs* e *gyrA* numa amostra de 44 isolados clínicos que inclui os 26 isolados XDR-TB analisados previamente. As mutações mais prevalentes encontradas nos genes pesquisados foram: Ins755GT, A1401G, G-10A e S91P nos genes *tlyA*, *rrs*, *eis* e *gyrA*, respectivamente. Foram ainda detectadas duas estirpes isoladas nos anos 90 que apresentariam já um perfil mutacional característico de XDR-TB. Estes dados são indicativos de que a XDR-TB em Portugal seria já uma realidade nos anos 90, na altura associada ao *cluster* Lisboa3.

Uma vez que a base molecular da resistência à estreptomicina era desconhecida em cerca de 33% dos isolados MDR com resistência à estreptomicina, foi investigado o possível papel de mutações ao nível do gene *gidB*. A pesquisa de mutações neste gene revelou a ocorrência de duas mutações não-sinónimas (*missense*), das quais uma – A80P foi encontrada somente em isolados sem mutações nos genes *rrs* e *rpsL* e exclusivamente associada ao *cluster* Q1. A caracterização do nível de resistência à estreptomicina, através de um teste de susceptibilidade a antibacilares semiquantitativo, mostrou que as estirpes com a mutação *gidB*A80P apresentavam um nível de resistência intermédio à estreptomicina. Os resultados aqui obtidos apontam portanto no sentido de a mutação A80P no gene *gidB* ser responsável por um nível resistência à estreptomicina mais reduzido e ainda poder ser utilizada como marcador filogenético para o *cluster* Q1.

Uma vez estabelecidos os principais mecanismos de resistência aos fármacos de primeira linha e segunda linha e, também avaliada a clonalidade da estirpes circulantes e responsáveis pela maioria dos casos de M/XDR-TB, procurou-se caracterizar melhor estas estirpes. Para tal, foi analisado um conjunto de 56 isolados clínicos de *Mycobacterium tuberculosis*, a maioria dos quais provenientes da Região de Saúde de Lisboa. Estas estirpes foram inicialmente genotipadas pela técnica de MIRU-VNTR com 24 *loci*. Verificou-se que o número adicional de *loci* face à mesma técnica mas utilizando apenas 12 *loci*, permitiu discriminar o *cluster* Lisboa3 em dois *clusters* distintos: Lisboa3-A e –B, sendo que apenas o *cluster* Lisboa3-B estava associado a casos de XDR-TB. Foi igualmente detetada a presença do *cluster* Q1, com associação a casos de XDR-TB, em que a utilização de *loci* adicionais permitiu a discriminação de alguns isolados clínicos, presumivelmente devido a fenómenos de evolução diversificante ao nível destes *loci.*

O genoma dos 56 isolados clínicos referidos foi igualmente sequenciado na totalidade através de uma plataforma de *Next Generation Sequencing* (Illumina HiSeq 2000). Os dados genómicos resultantes permitiram verificar uma elevada predominância de isolados pertencentes ao SCG 5 e, a construção de uma filogenia baseada em SNPs, na qual foram incluídas 19 estirpes de *Mycobacterium tuberculosis* cujo genoma já havia sido previamente publicado. Do ponto de vista filogenético, pôde observar-se que as estirpes Lisboa3 e Q1 constituem dois clades monofiléticos distintos mas bastante próximos. Os dados produzidos através desta abordagem contribuem para o conhecimento acerca da diversidade genómica de *Mycobacterium tuberculosis* tendo permitido identificar mutações adquiridas no processo de microevolução associado à aquisição de resistência aos antibacilares. Foram identificas possíveis mutações

compensatórias para a aquisição de resistência à rifampicina nos genes *rpoB* e *rpoC*, que codificam para diferentes subunidades da RNA polimerase.

Foram igualmente identificadas possíveis variantes genómicas estruturais específicas para diferentes clades filogenéticos aqui apresentados. De facto, foi identificada uma deleção de 112-3 bp no gene PPE41 nos isolados do clade Lisboa3, e é aqui proposta como um possível mecanismo de evasão imunitária, logo como uma vantagem seletiva destas estirpes. O mapeamento das sequências de inserção IS*6110* comprovou igualmente o importante papel destas sequências de inserção na génese de diversidade genómica funcional. Presume-se que estas sequências de inserção tenham um papel importante na evolução do Complexo *Mycobacterium tuberculosis* uma vez que a sua transposição afeta a integridade e regulação génica.

A análise dos rácios de mutações não-sinónimas/sinónimas revelaram ainda heterogeneidades ao nível do cromossoma bacteriano, genótipo e *Clusters* de Grupos Ortólogos, realçando possíveis e diferentes estratégias de evolução. Globalmente estes resultados apoiam a noção de uma crescente diversidade genómica face a uma adaptação geográfica e ao hospedeiro.

Palavras-chave: Mycobacterium, XDR, resistência, sequenciação genómica, MIRU-VNTR, epidemiologia molecular.

ABSTRACT

Multidrug-resistant (MDR) tuberculosis (TB) constitutes a major public health concern in Portugal. The present work was aimed at describing the MDR-TB strains circulating in Lisbon Health Region from a molecular viewpoint and, simultaneously assess its genomic diversity and phylogenetic origin.

Here, we start by describing the main mutations associated with first-line drug resistance and by revealing the presence of a high rate (56%) of extensively drug resistant (XDR) TB. Resistance to second-line drugs is further substantiated by the description of mutations found in second-line drug resistance associated genes and its association with genetic clusters, indicating clonal expansion in the community. Typing by Mycobacterial Interspersed Repetitive Unit – Variable Number of Tandem Repeats (MIRU-VNTR) revealed two major *clusters* of strains - Lisboa3 and Q1, mainly responsible for the M/XDR-TB strains in circulation. Review of laboratory data from a 6-year period confirmed the high and continued prevalence of XDR-TB strains, which was found to oscillate between 44.3-57.3% in the studied period.

Analysis of the *eis* gene promoter region identified a hypermorphic mutation (G-10A) in this region that is associated with kanamycin resistance in Lisboa3 isolates. Additionally, a mutational screening of the *gidB* gene revealed a missense mutation (A80P) that is proposed to confer intermediate-level streptomycin resistance and simultaneously act as a marker for Q1 *cluster* isolates.

Whole-genome sequencing of 56 Mycobacterium tuberculosis isolates allowed the construction of a genome-wide SNP-based phylogeny showing that Lisboa3 and Q1 isolates

form two distinct and unique monophyletic clades. Further genome analysis led to the identification of putative rifampicin resistance compensatory mutations and the identification of clade-specific genomic structural variants. In fact, a 112-3bp deletion found among Lisboa3 isolates is proposed to contribute towards immune evasion and virulence. IS*6110* is also addressed as a major driver in functional genomic diversification with putative impact on host adaptation.

Keywords: Mycobacterium, XDR, drug resistance, whole genome sequencing, MIRU-VNTR, molecular epidemiology.

ABBREVIATIONS

3R	Replication, recombination and repair
µg	Micrograms
µl	Microlitres
aa	aminoacid
AD	*Anno Domini*
AFB	Acid Fast Bacilli
AG	Arabinogalactan
aLRT	Approximate Likelihood Ratio Test
AMK	Amikacin
ATP	Adenosine triphosphate
ATS	American Trudeau Society
BC	Before Christ
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
CAP	Capreomycin
CAS	Central Asian
CDC	Centers for Disease Control and Prevention
CFUs	Colony Forming Unit
CFZ	Clofazimine
CIP	Ciprofloxacin
COGs	Cluster of Orthologous Groups
CS	Cycloserin
CTAB	cetyl trimethylammonium bromide
DOTS	Directly Observed Short-Course Treatment

DST	Drug Susceptibility Test
E. coli	*Escherichia coli*
EAI	East-African-Indian
ECDC	European Centre for Disease Prevention and Control
EEA	European Economic Area
ELISA	Enzyme-Linked Immunosorbent Assay
ELISpot	Enzyme-Linked Immunospot
EMB	Ethambutol
ETH	Ethionamide
EU	European Union
EU-15	Countries in the European Union prior to 1 May 2004
FASII	Type II Fatty Acid Synthase System
FQ	Fluoroquinolone
GTR	General Time Reversible
GTX	Gatifloxacin
GWAS	Genome-Wide Association Studies
h	Hours
HGT	Horizontal Gene Transfer
HIV	Human Immunodeficiency Virus
IFN	Interferon
IGRA	Interfer Gamma Release Assay
INH	Isoniazid
INSA-RJ	National Institute of Health Dr. Ricardo Jorge
IS	Insertion Sequence
KAN	Kanamycin
Kbp	Kilo base pairs
km^2	Square kilometers
L	Litre
LAM	Latin American Mediterranean
LAraM	Lipoarabinomannan
LJ	Löwenstein-Jensen
LSP	Large Sequence Polymorphism
LTBI	Latent Tuberculosis Infection
LTR	Long Terminal Repeats

M. africanum	*Mycobacterium africanum*
M. bovis	*Mycobacterium bovis*
M. canetti	*Mycobacterium canetti*
M. microti	*Mycobacterium microti*
M. tuberculosis	*Mycobacterium tuberculosis*
MDR	Multidrug Resistant/Multidrug Resistance
mg	Miligrams
MGIT	Mycobacterial Indicator Growth Tube
MIC	Minimum Inhibitory Concentration
min	Minutes
MIRU	Mycobacterial Interspersed Repetitive Unit
ml	Mililitres
MOX	Moxifloxacin
MTC	*Mycobacterium tuberculosis* Complex
NAAT	Nucleic Acid Amplification Test
NAD	Nicotinamide Adenine Dinucleotide
ng	Nanograms
NGS	Next-Generation Sequencing
N_s	Non-synonymous
nsSNP	Non-synonymous Single Nucleotide Polymorphism
nt	Nucleotide
ºC	degrees Celsius
OFX	Ofloxacin
ORF	Open Reading Frame
PAS	Para-mino Salicylic Acid
PCR	Polymerase Chain Reaction
PGG	Principal Genetic Group
PGM	Personal Genome Machine
PHA	Portuguese Health Authorities
pmol	Picomoles
POA	Pyrazinoic Acid
PPD	Purified Protein Derivative
PTH	Prothionamide
PZA	Pyrazinamide

Pzase	Pyrazinamidase
QBP	Quinolone Binding Pocket
qDST	semiquantitative Drug Susceptibility Testing
QFT-GIT	QuantiFERON®-TB Gold In-tube
RD	Region of Difference
REMA	Resazurin Microplate Assay
RFLP	Restriction Fragment Length Polymorphism
RIF	Rifampicin
RNA	Ribonucleic Acid
RRDR	Rifampicin Resistance Determining Region
rRNA	Ribossomal Ribonucleic Acid
S	Synonymous
SCG	SNP Cluster Group
sec	Seconds
SNP	Single Nucleotide Polymorphism
sSNP	Synonymous Single Nucleotide Polymorphism
STP	Streptomycin
SV	Structural Variant
TB	Tuberculosis
TNF	Tumor Necrosis Factor
T_s	Transition
T-SPOT	T-SPOT®.TB
TST	Tuberculin Skin Test
T_v	Transversion
UPGMA	Unweighted Pair Group Method with Arithmetic Averages
VCF	Variant Call Format
VIO	Viomycin
VNTR	Variable Number of Tandem Repeats
vs	*versus*
WGS	Whole-Genome Sequencing
WHO	World Health Organization
XDR	Extensive Drug Resistant/Extensive Drug Resistance

TABLE OF CONTENTS

TABLE INDEX

FIGURE INDEX

CHAPTER 1

General Introduction

1.1 BRIEF HISTORIC PERSPECTIVE

It is assumed that tuberculosis (TB) and mankind coexist since pre-historic times, having co-evolved through time. Human beings have emerged in Africa and formed small hunter-gatherer family groups, composed by one or two dozens of individuals. However, their dialectic groups could comprise a few hundred individuals and, numerous groups such as these could have had the capacity to withstand TB epidemics. When these ancestral humans migrated out of Africa, and even before settling in stable communities, they took this disease with them as they peopled the world. Afterwards, TB epidemic explosions took place as sedentary societies were settled.[1, 2]

The way human TB has appeared has been surrounded by some controversy, since the first theories stated that ancestral human TB would have emerged due to cattle domestication, leading to the contagion of ancestral human beings with the bovine tubercle bacilli. According to this theory, the modern human tubercle bacilli would have evolved from the bovine tubercle bacilli.[1] However, and since no domesticated cattle existed in North America until the arrival of the Spanish in the 15th century, it is more likely that human and bovine tubercle bacilli have evolved from a common ancestral. Recent molecular data point towards this latter hypothesis as the most likely.[3, 4]

The oldest Neolithic graves with TB signs date back to 5000 BC and 4000 BC, discovered near Heidelberg, Germany, and Liguria, Italy.[5] Both graves contained mortal remains evidencing the TB characteristic vertebral destruction. It is however in Ancient Egypt that the greatest number of mummified mortal remains, with vertebral lesions indicating TB, have been retrieved. Posterior analyses have detected the presence of DNA from the bacterial complex responsible for this disease (*Mycobacterium tuberculosis* Complex, MTC) in several mummies from a 3000 year time period (3400-500 BC). [6-10] In a study by Zink *et al*, MTC DNA was detected in 25 out of 85 Egyptian mummies, of which it was possible to genotype 12 by spoligotyping and exclude the bovine tubercle bacilli as the cause of the disease.[11]

In the American continent, TB was already present well before the arrival of Columbus in the 15th century. Mummies found in Peru (700 and 1000 AD) showed the same vertebral lesions characteristic of this form of the disease and, in one case MTC DNA was detected.[12, 13] Furthermore, several skeletons from American Indian graves were found with TB signs and there are examples of Mayan artwork and drawings that portray TB symptoms.[1]

Tuberculosis was also present in the pre-historic Asia and several references to a similar disease in India and China between 2000 and 1300 BC exist, with commentaries regarding symptomatology and disease characteristics already mentioning it as incurable.[1, 14]

There are Assyrian terracotta plates from the 7th century BC that describe patients coughing blood. The disease has been described by Hippocrates in the 4th century BC, as occurring mainly in individuals aged between 18 and 35 years old and used the term *phthisis*.[1, 15] Later, it was Galen that described TB in greater detail, suggesting that it was an infectious disease and even trying different treatments without any success.[1]

With the demographic explosion that took place in Europe during the second millennium AD and the enormous urban centre growth, Europe became the epicentre of innumerous epidemics, namely those that started in the 16th and 17th centuries.[15]

Between the 17th century, and until later in the 19th century, TB receives the name of white plague. The period that has registered the largest number of cases occurred in the second half of the 18th century, at a time that is estimated that 1.25% of the Europeans died of TB every year. England alone had an annual mortality rate of 1120 cases per 100 000 habitants due to TB. Another study from the early 19th century performed at the Hôpital de Charitê in Paris, revealed that from 696 examined corpses, 250 would have died of TB.[1, 15]

In the second half of the 19th century the TB associated mortality rate started to decline as sanitation and housing conditions improved (*e.g.* Paris urban renewal).

The cause for TB remained, nevertheless, unknown for centuries. In Northern Europe it was generally believed that it was a hereditary disease, while in Southern Europe it was regarded as an infectious disease. A first step in demonstrating that it was in fact an infectious disease was taken by Jean Antoine de Villemain, in 1865, by inoculating rabbits with human pulmonary extracts from tuberculous lesions and inducing similar lesions. However, only in 1882 it was definitely demonstrated by Robert Koch that TB was in fact an infectious disease, whose etiological agent is a bacteria, when he isolated and identified *Mycobacterium tuberculosis* (*M. tuberculosis*).[15, 16]

Still in late 19th and early 20th centuries, an experience conducted by Edward Trudeau showed that TB could be induced in rabbits using virulent cultures, with a variable prognosis depending on the environmental conditions in which the animals were kept in. This experience has greatly contributed to promote the sanatoria movement that had already started in Europe in late 19th century (Figure 1.1). The sanatoria were places in which patients were committed for several

months or years, with a treatment based on resting, fresh air and a balanced diet. Although today is still doubtful if this treatment improved the patient survival rate, the fact that these stayed isolated from the rest of the community may have effectively contributed for the reduction of the number of cases.[1, 17]

Figure 1.1 - Patients resting in reclining chairs in the BASF Sanatorium (1892) in Dannenfelds, Germany Extracted from Campbell.[18]

In early 20th century, some surgical techniques started to be used in TB treatment. Many of these techniques relied on the principle that the lung would need to rest in order to heal the lesions. Techniques such as the artificial pneumothorax that induced lung collapse and allowed the lung to rest started to be employed. More drastic were the thoracoplasty (surgical removal of ribs) or extra pleural plombage (paraffin infection or other inert material in the pleural cavity) that induced a permanent pulmonary collapse (Figure 1.2). Another procedure would consist in the unilateral crushing of the phrenic nerve, which would stop the diaphragm activity, also with the purpose of lung resting. Although for many patients, the collapse therapies may have saved their lives it showed very high risks and complications in the long-term.[1, 17]

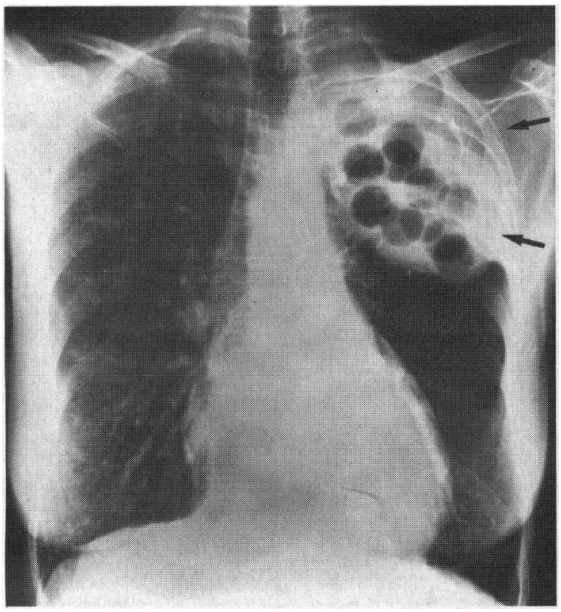


Figure 1.2 – Posteroanterior chest x-ray film from a 64-year old woman treated for tuberculosis by stripping the parietal pleura from the chest wall and packing the space with inert Lucite balls (arrows). Extracted from Mond & Khan.[19]

In the 1920s, Albert Calmette and Camille Guerin produce the first vaccine for TB prophylaxis, named Bacillus Calmette-Guerin (BCG), from the sequential passage of a virulent *Mycobacterium bovis* (*M. bovis*) strain. Later, in the 40s the era of TB antibiotic treatment starts with Albert Schatz and Selman Waksman and the discovery of streptomycin (STP). In the

following years the main antibacillary drugs still used today to fight TB are (re)discovered, *e.g.*, isoniazid (INH), rifampicin (RIF) and pyrazinamide (PZA). [1, 17]

The implementation of adequate treatment regimens with effective antibacillary drugs has lead to an accentuated reduction in the number of cases in many countries during the 60s and 70s, which also lead to the belief that the fight against TB was won. Nevertheless, the combined emergence of resistant strains and the human immunodeficiency virus (HIV) are still today threatening the control of this disease. [1, 17]

1.2 TUBERCULOSIS IN THE PRESENT

1.2.1 Tuberculosis in the World

The worldwide number of TB cases still reaches alarming proportions. Although at global level the disease incidence and prevalence present a decreasing trend in last 10 and 15 years, respectively, the most recent report from the World Health Organization (WHO) point out to an estimated occurrence of 8.7 million new cases in 2011 (Table 1.1). This number of incident cases is reflected by a high global incidence rate of 125 cases per 100 000 habitants. The most affected regions are Asia and Africa, with 59 and 26% of the total number of new cases worldwide. On the other hand, countries like India in which the total number of new cases (2.2 million) corresponds to 26% of the total number of new cases worldwide, simultaneously show a comparatively lower incidence rate (181 cases per 100 000 habitants). Other countries, such as South Africa and Zimbabwe, due to the high HIV co-infection rates have incidence rates of 993 and 603 new cases per 100 000 habitants, respectively.[20]

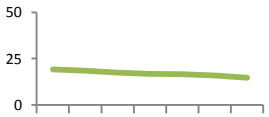
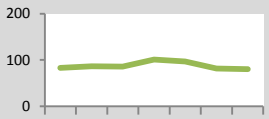
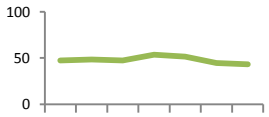
Regarding mortality due to TB, WHO estimates that 1.4 milion deaths have occurred in 2011 due to TB, of which 0.43 million took place among HIV-positive individuals. Although number of deaths stratified by region was only reported by the WHO for HIV-negative patients: in absolute numbers, it was verified a higher number of mortal cases in the Southeast Asia region with an estimate of 480 000 deaths. Mortality rate was higher in the African and Southeast Asia regions, both with a mortality rate of 26 deaths per 100 000 habitants (Table 1.1).[20]

Table 1.1 – Incidence, prevalence, mortality and HIV coinfection rate across the different WHO Regions in 2011. Data extracted from the WHO surveillance report.[20]

WHO Region	Incidence		Prevalence		Mortality[a]		HIV
	thousands	per 100 000 Hab.	thousands	per 100 000 Hab.	thousands	per 100 000 Hab.	%
Africa	2 300	262	2 500	293	220	26	39.0
Americas	260	28	330	35	21	2.2	14.0
Eastern Mediterranean	660	109	1000	170	99	16	1.5
Europe	380	42	500	56	45	5.0	6.1
Southeast Asia	3 500	189	5000	271	480	26	4.1
Western Pacific	1 700	92	2500	138	130	6.9	2.2
Global	8 700	125	12 000	170	990	15	13

[a]WHO Mortality indicator due to TB excludes deaths among HIV-positive patients

Table 1.2 – TB notification rates (per 100 000 habitants) and tendency across the European Region between 2004-2010. Data extracted from the European Centre for Disease Prevention and Control/WHO Regional Office for Europe surveillance report.[21]

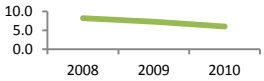
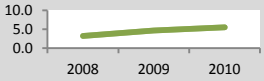
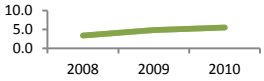
Countries	2004	2005	2006	2007	2008	2009	2010	Trend
EU/EEA	19.1	18.5	17.5	16.8	16.5	15.8	14.6	
Non-EU/EEA	82.8	86.5	85.5	100.6	96.7	81.3	80.1	
Europe	47.3	48.2	47.3	53.4	51.6	44.4	43.2	

In Europe, according to the European Centre for Disease Prevention and Control (ECDC), 388 875 new cases were notified in 2010 resulting in an apparent decline of 15.2% in the period of 2005-2010. Notification rates registered for each country in this region ranged between 2.8 and 123 cases per 100 000 habitants in Italy and Kazakhstan, respectively. In 2010, the global notification rate for the European region was 43.2 cases per 100 000 habitants, although it was significantly lower in the European Union (EU) and European Economic Area (EEA) countries than in the other countries group: 14.6 *vs* 80.1 cases per 100 000 habitants, respectively (Table

1.2). Both country groups have presented a decreasing trend in the notification rates, more accentuated in the EU/EEA countries, -4.4% in the 2006-2010 period, when compared with the other countries in the same period (-0.9%) (Table 1.2).[21]

The HIV/TB co-infection rate globally registered in Europe was 5.5% in 2010, with a growing tendency in the last 3 years. This tendency is the result of diverging tendencies registered among the groups of EU/EEA and other countries. In the EU/EEA countries a decreasing tendency in the number of TB/HIV co-infected cases has been registered in the last 3 years, while the remaining countries have registered an increase in the TB/HIV co-infection rate in the same period that results in the increasing global European co-infection rate (Table 1.3).[21]

Table 1.3 – Number, percentage and tendency of TB/HIV co-infection between 2008-2010. Data extracted from the European Centre for Disease Prevention and Control/WHO Regional Office for Europe surveillance report.[21]

Countries	2008		2009		2010		Trend (%)
	N	(%)	N	(%)	N	(%)	
EU/EEA	1402	8.2	1334	7.3	1055	6.0	
Non-EU/EEA	10397	3.2	12187	4.6	14779	5.5	
Europe	11799	3.4	13521	4.8	15834	5.5	

With the exception of Ukraine, all countries that have a TB/HIV co-infection rate above 10% belong to EU/EEA. The six countries with the highest TB/HIV co-infection rates are: Ireland (17.6%), Portugal (13.3%), Ukraine (13.0%), Netherlands (12.7%), Estonia (11.5%) and Malta (11.5%).[21]

Concerning mortality in Europe due to TB, the ECDC estimates that 60 135 deaths occurred due to TB (6.7 per 100 000 habitants) in 2010 in the WHO European space. About 91% of these mortal cases occurred in countries other than the ones in the EU/EEA (54 752 deaths, 140 deaths per 100 000 habitants), in comparison with the countries that integrate this latter group (5 383 deaths, 1.1 deaths per 100 000 habitants).[21]

1.2.2 Tuberculosis in Portugal

Portugal is the western European country with the highest incidence rate with 2 626 cases in 2010 of which 2 398 are new cases, *i.e.*, with no previous treatment. This incidence results in a notification rate of 24.7 cases per 100 000 habitants or 22.5 cases per 100 000 habitants if only the new cases are considered.[21] This incidence rate can be classified as intermediate by comparison with the remaining European countries. According to the Portuguese Health Authorities (PHA) this notification rate has a decreasing tendency for the last 19 years with a mean annual reduction of 6.4%.[22]

The geographical distribution of the cases at a nationwide level show a higher case incidence in the districts of Lisbon and Oporto, the only districts with notification rates above 30 cases per 100 000 habitants. Lisbon is according to the PHA data for 2010 the district with highest notification rate (31.5 cases per 100 000 habitants) and absolute number of cases (706 new cases).[22] Six districts, however, still present notification rates with values comprehended between 20-30 cases per 100 000 habitants: Viana do Castelo, Vila Real, Bragança, Setúbal and Faro. Provisory PHA data show that in 2011 the Oporto district had the highest TB notification rate (31.9 cases per 100 000 habitants).[23]

At a demographic level, a higher incidence of the disease was observed in male individuals at a 1.9:1 ratio and, a median age shift from the 25-34 age group to the 35-44 age group, occurred in the last decade.[22]

The number of cases observed among immigrants was 397 in 2010, 16% of the total notified cases. The origin of the majority of these foreign individuals was: Angola (23%), Guinea-Bissau (16%), Cape Verde (13%), Brazil (12%) and Mozambique (9%).[22] The 2011 PHA provisory data, showed 372 new cases among immigrants were observed with a significant increase in patients from Cape Verde (18%) and from Romania (7%).[23]

According to the ECDC, concerning drug resistant cases, 68 INH resistant and 19 RIF resistant cases were notified in 2010, with these latter also exhibiting INH resistant, *i.e.*, multidrug resistant (MDR).[21] In 2010, no case of extensively drug resistant (XDR) TB (multidrug resistance plus resistance to a fluoroquinolone (FQ) and an injectable second-line drug: kanamycin (KAN), amikacin (AMK) or capreomycin (CAP)) was notified.[21] PHA data, slightly differ from the ECDC data since the latter reports 29 cases of MDR-TB and 2 cases of XDR-TB nationwide (Figure 1.3).[22, 23] These numbers appear to suffer from some degree of under notification since they

are low when compared to the number of clinical isolates reported in peer-reviewed scientific publications in the last 15 years.[24-26] Also important to stress is that in the ECDC reported data only 61.1% of the notified cases are laboratory-confirmed and from these, drug susceptibility testing data was available for only 76.3% in a total of 1 225 cases.[21] For 2011, data from PHA show a marked increase in the number of XDR-TB cases.[23]

The indicators for treatment success for the 2009 cohort, 24 months after diagnosis, show a success rate of 84.2%.[22] For this cohort the following was observed: 88 deaths (5.6%), 48 treatment defaults (3.1%), 79 patients still on treatment after 24 months (5.0%) and 32 patients lost to follow-up (2.0%).[21] Analysis of the 2010 cohort show a reduction in the success rate to 77%.[23]

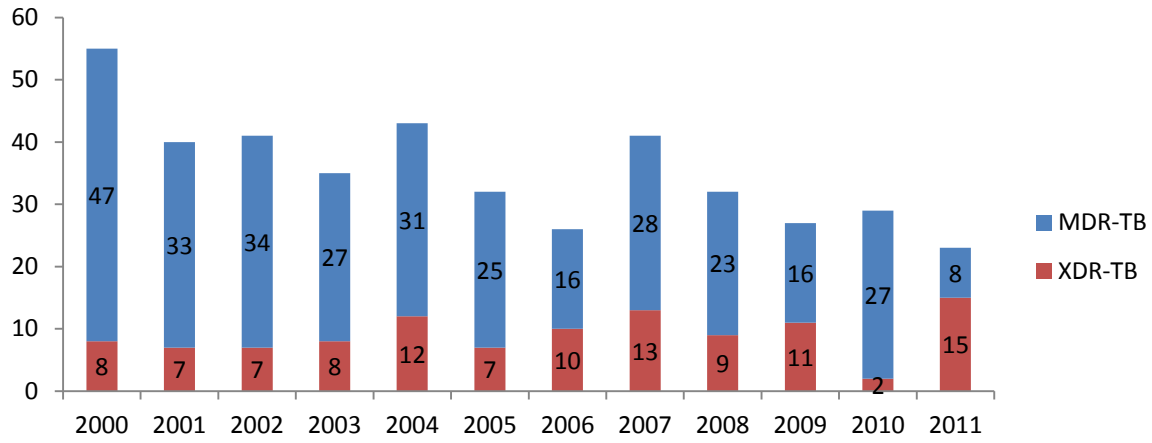


Figure 1.3 – Number of new MDR- and XDR-TB cases reported by PHA between 2000-2011. Data extracted from PHA reports.[22, 23]

1.3 MYCOBACTERIA

1.3.1 General Morphological Properties

Mycobacteria are slighly curved or straight bacillii, whose dimensions range between 0.2-0.6 x 1.0-10.0 µm, although some degree of pleomorphism may be exhibited. These are aerobic microorganisms that may display a ramified growth or a filamentous growth similar to a mycelium. Contrary to the remaining Actinomycetes, a true mycelium is not formed and when subjected to some physical disturbance, fragmentation in bacilli or cocci occurs.[27] Nevertheless, *Mycobacterium farcinogenes* and *Mycobacterium senegalense* do form a stable mycelium.[28] Although controversial, recent studies performed with *Mycobacterium marinum*

and *Mycobacterium avium* Subsp. Paratuberculosis seem to demonstrate the ability of some mycobacteria to produce endospores as a possible mechanism of persistence in the host.[29-32]

Regarding the colonial morphology, mycobacteria may present more than one type of colony, even varying with the species' strain. Usually, *M. tuberculosis* strains present rough colonies while the colonies of *Mycobacterium intracellulare* are usually smooth. In other species the colony morphology depends on the growth phase.[28]

Mycobacteria may be strictly parasitic, saprophytes or exhibit an intermediate lifestyle. Most species free lives in the water or soil, but for many species, their ecologic niche are the tissues of homeothermic hosts, *e.g. M. tuberculosis*, whose natural reservoir is the human population.[28]

Mycobacterial cell wall has several particularities making it significantly different from the typical Gram positive and negative cell wall types. The mycobacterial peptidoglycan consists in glycosidic alternate chains composed by N-acetylglucosamine and muramic acid binded by a β(1→4) bond. The tetrapeptid, essential for cross-linking with other glycosidic chains is bound to the muramic acid. This tetrapeptid consists in L-alanil-D-isoglutaminil-meso-diaminopelil-D-alanin with an amide on the diaminopimelic acid residue. The muramic acid is acylated with a glycolyl residue instead of an acetyl residue and, the number of cross-links between the peptidoglycan chains are higher than in the peptidoglycan from *Escherichia coli*.[34]

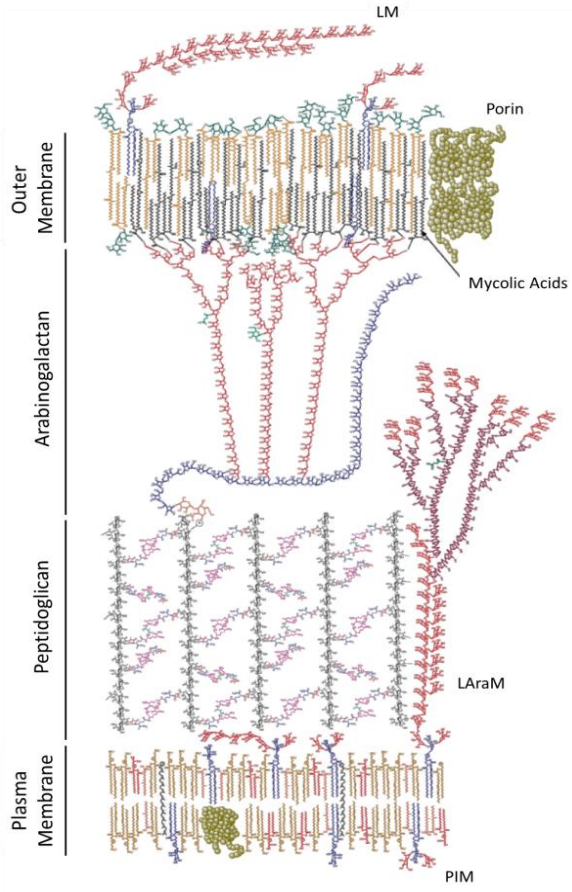


Figure 1.4 – Theoretical model of the *M. tuberculosis* cell envelope (composed by the plasma membrane, cell wall core and outer membrane) chemical structure showing its major components. Abbreviated molecules: LAraM, lipoarabinomannan; LM, lipomannan; PIM, phosphatydil-*myo*-inositol mannosides. Adapted from Kaur *et al*.[33]

Another molecule, the arabionogalactans (AG), are covalently bound to the peptidoglycan. These molecules have unusual monosaccharides and bonds and, are composed by D-arabino-furanosyl and D-galacto-furanosyl.[34] The AGs are esterified by mycolic

acids, of great importance in the taxonomical classification of mycobacteria, which are long-chained α-ramified-β-hidroxylated fatty acids that may contain up to 90 carbon atoms.[28, 34] The presence of mycolic acids confers on mycobacteria the property of acid-fastness, at least during some stage of its life cycle. Although mycobacteria are not stainable by Gram staining method, due to the lipidic content on the surface, if the lipidic portion is removed with alkaline ethanol, the reminescent cell becomes gram-positive. [35]

A complex mixture of proteins and polysaccharides has also been described around *M. tuberculosis in vitro* when non-mechanically disturbed. This structure is possibly a pseudocapsule since it is not covalently bound to the cell wall and is composed by polysaccharides such as glucans, arabinomanans, manans and proteins. Part of the proteins that compose this pseudocapsule appear to be proteins that are being transported to the exterior while others are probably permanent proteins. Besides being present in culture filtrates, these proteins are also present in considerable quantities in cell extracts obtained by mechanical treatment.[34, 36]

The generation time is highly variable from species to species, oscillating between 2-20 hours, which implies that colony formation may take from 2 days to 8 weeks when incubation is performed at the optimal temperature. The optimal temperature also varies according to the species, between 30-45ºC.[28]

Culture may be performed in simple media with mineral salts, using ammonia or amino acids as nitrogen source and glycerol as a carbon source. There are, nevertheless, species that require supplementation such as hemin, micobactins or other iron transporter.[28]

Many mycobacteria produce yellow or orange carotenoid pigments, whose induction may or not occur by photo induction. The presence of these carotenoids may protect against oxidative damage by oxygen in the singlet form.[27]

Mycobacteria also have a high GC content (>50%), that may range between 62-70%.[28]

1.3.2 Mycobacterial taxonomy

Although mycobacteria were one of the early described bacterial groups, a functional taxonomy was only elaborated in the last 50 years. Mycobacteria are included in the *Mycobacteriaceae* family, that contains a single genus – *Mycobacterium*, with a present

number of 156 species and 13 subspecies.[37, 38] This number has increased significantly in the last two decades.[39, 40]

Phylogenetically close are the *Corynebacterium*, *Nocardia* and *Rhodococcus* genera, which is reflected by their high GC content and, still, by the fact that these also produce mycolic acids, although mycobacterial mycolic acids contain a higher number of carbon atoms.

A natural division exists between fast-growth and slow-growth mycobacteria. Fast-growth mycobacteria take less than seven days to develop colonies from a diluted inoculum, while fast-growers take more than seven days to form visible colonies.

Tsukamura has proposed, in 1967, the separation of fast-growers and slow-growers in two distinct sub-genera. This proposal was not adopted since both groups demonstrate a considerable degree of similarity regarding DNA homology, antigenic and lipidic composition, and antibiotic susceptibility.[28]

Other mycobacterial classification in three groups can be made based on pigmentation:

- Non chromogenic – do not exhibit pigment production, *e.g.*, *M. tuberculosis* or *M. bovis*;
- Photochromogenic – produce pigments through photo induction, *e.g.*, *M. marinum* e *M. kansasii*;
- Scotocromogenic – produce pigments in the absence of light, *e.g.*, *M. gordonae*.

In 1959, Runyon proposed the classification of mycobacteria in 4 groups, based on pigmentation and growth rate:

- Group I – photocromogenic slow-growth mycobacteria;
- Group II – scotocromogenic slow-growth mycobacteria;
- Group III – non-cromogenic slow-growth mycobacteria;
- Group IV – fast-growth mycobacteria.

Although this is a phenotypic-based classification and does not reflect phylogenetic proximity, the Runyon classification is still used today. [41]

Stahl *et al* have verified through 16S rRNA comparison that the inferred phylogeny reflects the division between slow-growth and fast-growth mycobacteria, with a greater coherence between the slow-growers as these formed a distinct line of evolutionary descendance (Figure

1.5).[42] It is however difficult to discriminate all mycobacteria at the species level since closer species can have a 16S rRNA similarity greater than 97%, the generally accepted limit for the inclusion of strains in the species taxon. [27, 43]

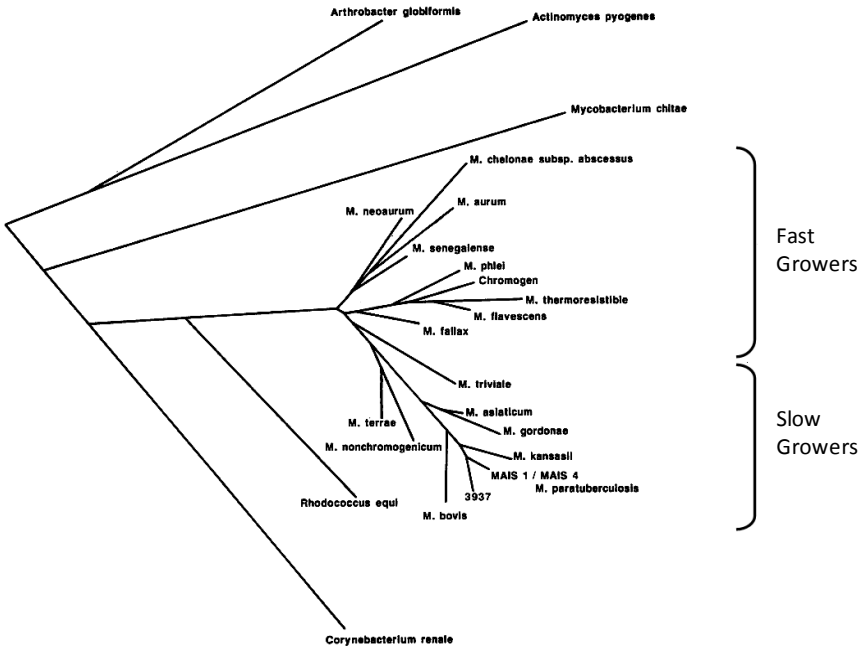


Figure 1.5 – 16S rRNA-based phylogeny of several mycobacterial species and related microorganisms. Adapted from Stahl *et al*.[42]

More recent multigenic studies, involving not only the nucleotide sequence of the 16S rRNA but also nucleotide and amino acidic sequences of additional genes (*e.g.*, *hsp65* and *rpoB*) have allowed a better discrimination of mycobacterial species and confirm the same phylogenetic division and monophyletic origin of slow-growth mycobacteria.[43-45] According to this phylogeny, the slow-growers' group has probably evolved from one of the two main fast-growth mycobacterial groups.[44, 45]

1.3.3 *Mycobacterium tuberculosis* complex

TB-causing mycobacteria form a group designated by *Mycobacterium tuberculosis* complex (MTC), that includes *M. tuberculosis*, *M. bovis* (subspecies *bovis* and BCG), *Mycobacterium africanum* (*M. africanum,* subtypes I and II), *Mycobacterium microti* (*M. microti*) and M*ycobacterium canetti* (*M. canetti,* not officially recognized). The MTC species can be considered as ecotypes of the same species, adapted to different hosts. Nevertheless, since the MTC species have been formally described, the above nomenclature is still used.

In 2003, Cousins *et al* have proposed the integration of a new member in this complex, *Mycobacterium pinnipedii,* that has already been characterized and integrated in the MTC.[46] In the same year, it was equally recognized as belonging to the MTC the *Mycobacterium caprae* species, initially described as a *M. tuberculosis* subspecies and later as a *M. bovis* subspecies.[47-49] More recently, two new species were proposed to be added to the MTC, *Mycobacterium mungi* and *Mycobacterium orygis*.[50, 51]

Genetically, all members of this complex are quite similar, having about 99.9% of nucleotide sequence similarity and identical 16S rRNA sequences.

MTC species can be differentiated be phylogenetic-informative genomic deletions that seem to demonstrate that *M. microti*, *M. bovis* and *M. africanum* have a more recent common ancestor than with *M. tuberculosis*. This phylogeny suggests that the ancestral species of *M. tuberculosis* has diverged, which was followed by speciation events that have lead to the origin of the remaining MTC species, with the exception of *M. canetti*. Brosch *et al* also proposed that the common ancestor for the entire MTC would be similar to *M. tuberculosis* or *M. canetti* (the first species to diverge) and would already be a human pathogen.[3]

1.3.4 Genetic Diversity, Evolution and Adaptation

For several years, the *M. tuberculosis* species has been regarded as a highly specific adaptation to the human species, considered to be its only reservoir, possibly derived from another species with a broader host variety, such as *M. bovis*. Sreevatsan *et al* have classified *M. tuberculosis* as a recent species and proposed that the speciation event would have occurred between 25 000 – 30 000 years ago. According to these authors, the fact that it was a fairly recent species would explain the reduced genomic variability found when characterizing 26 structural genes from clinical isolates.[52]

More recent comparative genomic analysis has lead to the conclusion that the between-lineage existing genetic diversity is higher than the one initially estimated. Gutierrez *et al* when comparing the accumulated differences in six structural genes of 17 strains have estimated that the speciation would have occurred between 2.6 – 2.8 million years ago.[53] This estimate is therefore higher than the proposed by Sreevatsan *et al* and places *M. tuberculosis* as one of the most ancient pathogens, that has inclusively affected the early hominids.[52, 53]

During the last two decades several techniques have been employed in the assessment of *M. tuberculosis* diversity, some with a high discriminatory power and, hence, of enormous utility in molecular epidemiology through the identification of outbreaks at a regional level (*e.g.*, RFLP-IS*6110* and MIRU-VNTR). Other techniques, such as single nucleotide polymorphism (SNP) or deletion typing, have a lower discriminatory power but, are able to position a given strain in a species' global population structure.

Spoligotyping, which is discussed ahead as a molecular epidemiologic technique (see Molecular Epidemiology section) is also extremely useful and one of the most widely diffused techniques to assess global strain diversity. Global data from the international spoligotype database (SITVIT) has allowed the definition of the major lineages within *M. tuberculosis* sensu stricto: Beijing, Central Asian (CAS), East-African-Indian (EAI), Haarlem, Latin-American-Mediterranean (LAM), S, T and X (Figure 1.6). Furthermore, a total of 62 MTC sub-lineages can be defined in the SpolDB4 version.[54-59] The latest version of SITVIT (SITVIT1) houses spoligotyping data from 62 582 clinical isolates and defines new rules for clade assignment.[60]

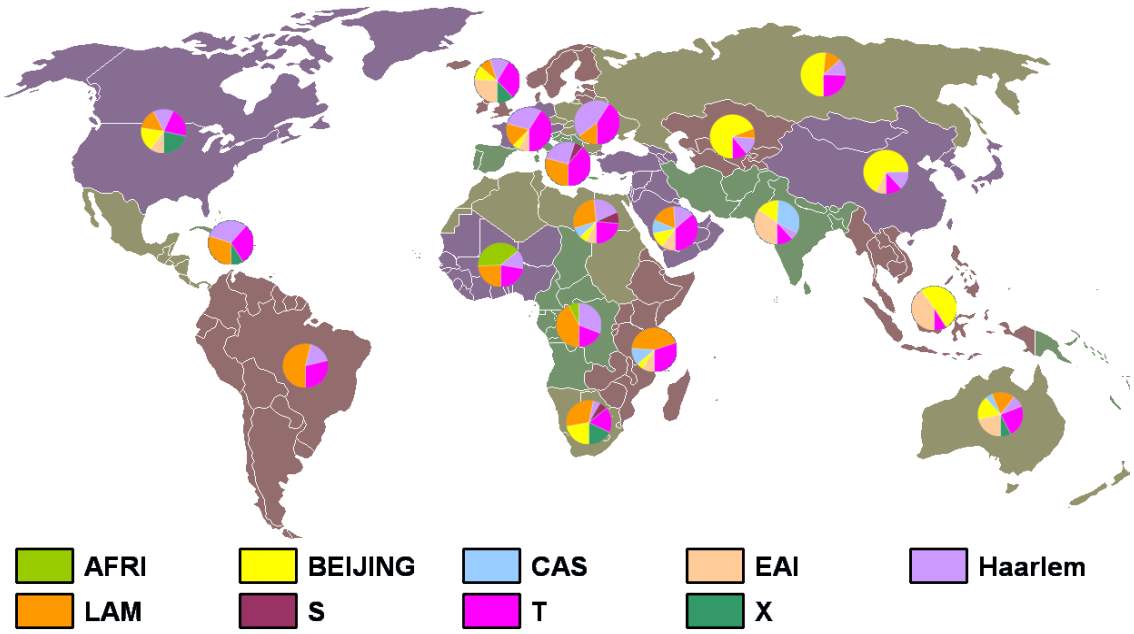


Figure 1.6 – Prevalence and distribution of the major spoligotype lineages. Extracted from the SITVITWEB database website.[60]

The development of further markers was greatly boosted by the knowledge of *M. tuberculosis* genome. This first genomic analysis was initially carried out by sequencing the genome of *M.*

tuberculosis H37Rv, a laboratory strain derived from a clinical isolate dated back to 1905, that has maintained its virulence on animal models.[61, 62] Presently, several genomes of *M. tuberculosis* clinical isolates have been made available, showing an average genomic length of 4.4 Mb, with a GC content averaging 65.4%, with around 4081 genes coding for 3970 proteins.

The genomic comparison of clinical isolates from different origins with the genomic sequence of *M. tuberculosis* H37Rv has allowed the identification of a considerable number of SNPs and the construction of a global phylogeny of different *M. tuberculosis* lineages. Four studies have tried to define the *M. tuberculosis* populational structure, with the proposed phylogenies presenting a high degree of congruence.[62-65] Sreevatsan *et al* when analysing 26 structural genes have identified two non-synonymous SNPs (nsSNPs) that have allowed the definition of three *M. tuberculosis* Principal Genetic Groups (PGG).[52] This division is also congruent with other proposed phylogenies based on higher numbers of SNPs. Filliol *et al* have analysed 212 isolates, by comparison of 159 synonymous SNPs (sSNPs), and defined six SNP cluster groups (SCG) and five sub-groups (Figure 1.7). Furthermore, this phylogenetic classification has allowed to define a restrict number of six SNPs that allows the classification in one of the SCGs.[64] Alland *et al*, on the other hand, defined a different set of nine SNPs that allows not only the SCG classification of a given strain, but also the sub-group identification.[66]

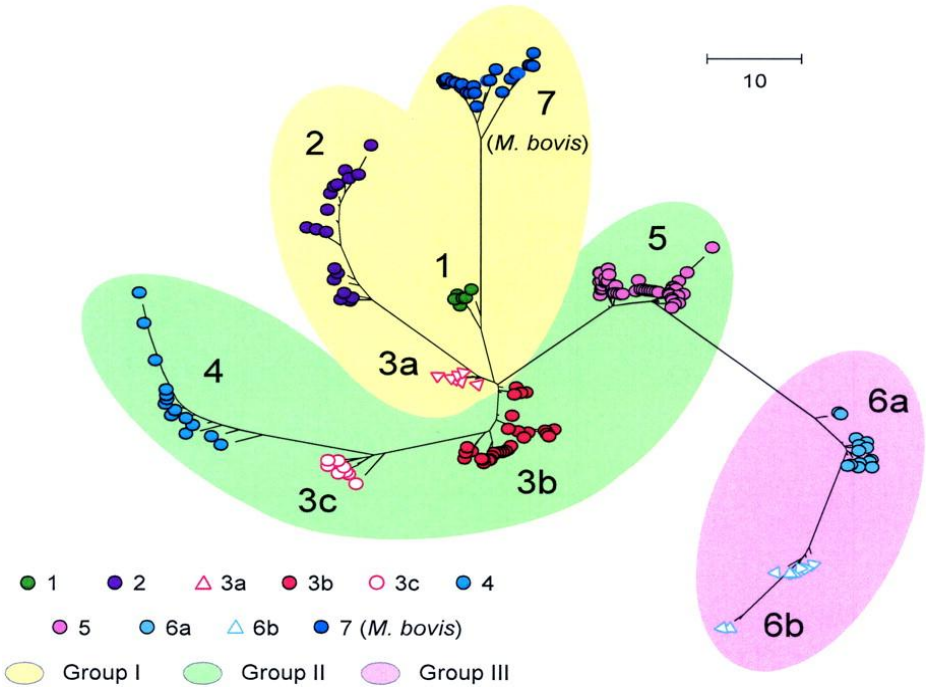


Figure 1.7 – Global phylogenetic tree of *M. tuberculosis* isolates. SCGs and its subgroups are highlighted with the respective designation and colour. PGGs are also highlighted by background colouring according with Sreevatsan *et al*. Scale bar indicates the phylogenetic distance in SNP number. Extracted from Filliol *et al*.[64]

Molecular data has shown that particular groups of strains with unknown selective advantages are the cause of MDR-TB outbreaks. The Beijing/W strains or the *M. bovis* B strains, in Spain, are good examples.[67-70] Their vast distribution reflects their success, and several studies have demonstrated that the Beijing/W strains induce a more severe pathology, suggesting an increased virulence.[71]

Phylogenetic classification is however insufficient for explaining possible selective advantages that might be behind high prevalence rates by some lineages or families between different human populations. However, other genetic alterations occurring in these lineages, with marked impact on strain physiology and adaptability, might be responsible for these differential prevalence rates.

Hershberg *et al* have concluded, in a study involving a collection of globally representative isolates of *M. tuberculosis* and other MTC members, that the MTC species are under a reduced purifying selection; MTC diversity is associated with human demographic and migratory events; and, that functional diversification is lead by genetic drift.[2]

Horizontal Gene Transfer (HGT) is a major driver in bacterial evolution and is thought to have occurred in the MTC ancestral species, *Mycobacterium prototuberculosis*, before the emergence of the remaining MTC members.[53, 72] HGT has probably played an important role in the transfer of pathogenicity islands, from other environmental bacteria, that may have conferred a selective advantage in colonization of protozoa.[72] Rosas-Magallanes *et al* have shown that an operon essential for successful macrophage infection has probably been acquired through HGT by the MTC ancestor from a γ-proteobacterium donor.[73] Furthermore, Becq *et al* in a genome-wide comparative study of *M. tuberculosis* H37Rv, CDC1551 and *M. bovis* AF2122/97, detected 48 genomic islands probably acquired by HGT and, found that 8.2% of the genes present in these are necessary for *in vivo* virulence.[74, 75] In this view, it was proposed that current MTC members appear to no longer acquire new genomic regions by HGT, or at least do it at an extremely low rate, probably in part due to their restricted ecological niche.[72, 76] A recent publication, using high-throughput sequencing data from *M. tuberculosis* clinical isolates and comparison with other mycobacterial genomes suggests that recombination is still occurring in contemporary MTC members, contributing to genomic diversification.[77]

In the MTC organisms, the main mechanism of genomic diversification is probably through large genomic deletions as it has been shown that the MTC species have evolved through genomic downsizing.[3, 78-81] Some of these deletions are considered to be phylogenetically

informative, allowing the distinction of different species belonging to the MTC and differentiation of isolates from *M. tuberculosis sensu stricto* and its evolutionary history.[3, 82] The phylogeny obtained by Gagneux *et al* is also congruent with the SNP-based phylogeny and possibly reflects better the first human migrations.[2, 78] Based on phylogenetically informative deleted regions, Large Sequence Polymorphisms (LSP) or Regions of Difference (RD), these authors have defined six main lineages, two of which include the species traditionally known as *M. africanum* (Figure 1.8). Another study aimed at defining global phylogeography through 12 *loci* Mycobacterial Interspersed Repetitive Unit - Variable Number of Tandem Repeats (MIRU-VNTR) defines three major phylogenetic lineages within *M. tuberculosis sensu stricto*, congruent with Gagneux's LSP classification.[83] An integration of the different genetic marker based classifications is presented in Table 1.4.

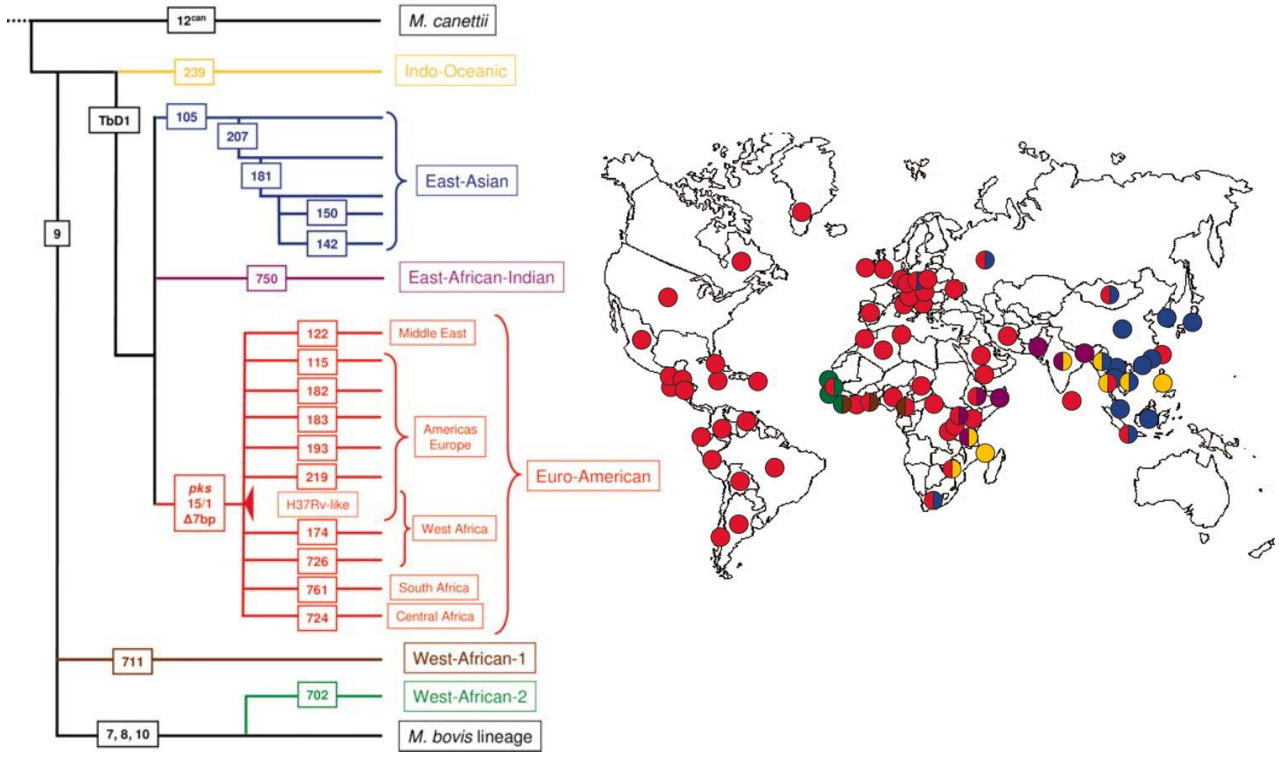


Figure 1.8 – RD-based phylogeny for different MTC lineages with greater resolution within *M. tuberculosis*. Branch rectangles indicate the clade-specific RDs. Also shown, is the geographical distribution of each lineage. Adapted from Gagneux *et al*.[78]

Table 1.4 – Integration of different phylogenetic nomenclatures based on different markers and geographical association. Adapted from Gagneux & Small.[84]

Genetic Marker	Lineage 1	Lineage 2	Lineage 3	Lineage 4	Lineage 5	Lineage 6
SNP Sreevatsan *et al* [52]	PGG1	PGG1	PGG1	PGG2-3	PGG1	PGG1
LSP Gagneux *et al* [78]	Indo-Oceanic lineage	East Asian lineage	East African-Indian lineage	Euro-American lineage	West African lineage I	West African lineage II
SNP Filliol *et al*[64]	SCG 1	SCG 2	SCG 3a	SCG 3b-6b	Not done	Not done
Spoligotyping Brudey *et al* [55] Demay *et al* [60]	EAI	Beijing	CAS	Haarlem, LAM, T, X	AFRI2, AFRI3	AFRI1
MIRU-VNTR Hill *et al* [83]	Indo-Oceanic	East Asian and African Indian	East Asian and African Indian	Euro-American	West African I	West African II
Geographical Association	East Africa, Southeast Asia, South India	East Asia, Russia, South Africa	East Africa, North India	Americas, Europe, North Africa, Middle East	Ghana, Benin, Nigeria, Cameroon	Senegal, Guinea-Bissau, The Gambia

Besides the existence of phylogenetic informative deletions, some deletions have been found to be acquired independently in different lineages (homoplasies).[66] These deletions probably represent polymorphisms with phenotypic impact and might occur and be selected due to selective pressures favouring that specific impacted phenotype. Newton *et al*, *e.g.*, have identified an Asian lineage defined by a deletion later associated with immune subversion.[85]

Other factors such as polymorphisms in recombination, replication and repair genes (3R) may account for the species' diversification and adaptation. Mutations in these genes, that assure the genome's stability, might lead to a mutator phenotype and contribute to the development

of mutations in other genes at an increased frequency. It is possible that this phenotype might contribute to an improved adaptation to environmental changes, *e.g.* anti-TB therapy.[86]

Moreover, the location of different insertion sequences, such as IS*6110*, might also affect the expression and integrity of specific genes.[87] This insertion sequence has been widely used as a molecular marker in molecular epidemiology given its highly variable nature, in number and differential position in the genome. Soto *et al*, when analysing the insertion site of IS*6110* of the MDR *M. bovis* B strain have verified that the insertion of an IS*6110* in the promoter region of the *phoP* gene was responsible for the overexpression of this virulence associated gene.[88] In a similar way, Otal *et al* have also verified that in *M. bovis* isolates, the differential location of several IS*6110* might constitute an adaptation factor to the host.[89] Safi *et al* further demonstrated that when *M. tuberculosis* grows inside human monocytes, the IS*6110* effectively acts as a promoter of upstream genes.[90]

Other polymorphisms found throughout the genome of different strains might provide comparative information essential to understand and gain insights on the basis of human colonization, infectiousness and virulence.[91] Hirsh *et al* have observed that an association exists between the different *M. tuberculosis* strains and the human populations to which they are genetically related. Such association appear to result from a co-adaptation or interaction between the human and bacterial genomes.[76]

Genomic and proteomic variations between strains can induce a more virulent phenotype or elicit different immune responses by the host.[92] Reed *et al* verified that a phenol glycolipid, produced by a subset of *M. tuberculosis* isolates belonging to the Beijing/W family, inhibits the release of key pro-inflammatory cytokines by the cells of the host's innate immune response.[93, 94] Manca *et al* found on the other hand, that *M. tuberculosis* CDC1551 induced in vitro and in vivo a more vigorous immune response.[95] More recently, Nahid *et al* showed that Euro-American lineage, sub-lineage RD724, was associated with more severe disease at baseline, and together with the East-Asian lineage was associated with lower bacteriological conversion after 8 weeks of treatment.[96] It becomes clear that genetically different strains can evoke different immunopathological events.[97] Further data on the genomic diversity and association with disease presentation and/or virulence, including early reports on differential virulence between different geographically isolated strains, has been extensively reviewed by Coscolla & Gagneux.[98]

1.4 TUBERCULOSIS INFECTION

1.4.1 Symptomatology and Mode of Transmission

TB as a disease can show a symptomatology that can vary depending on the infected individual or the affected organ. From a clinical perspective, primary TB is usually asymptomatic, but can, however, evolve to a milliary or post-primary TB if the immune response is not enough to control the initial infection.[99, 100]

The general symptoms of respiratory active TB generally include tiredness, fever, feeling of discomfort and anorexia. The weight loss occurs along with the progression of the disease, accompanied by sweating and fever, particularly during the night. Fever is usually low but chronic. Coughing is the most common respiratory symptom, being persistent (> 3 weeks) with mucoid or purulent sputum and possibly accompanied by hemoptysis (coughing sputum with blood). As mentioned, some degree of variability in the presentation of these symptoms may occur with individuals that show less symptoms.[99, 100]

Besides the respiratory system, TB can affect any organ and in such cases it is designed as extra-pulmonary or non-respiratory TB. The symptoms vary accordingly with the affected organ, *e.g.*, infection of the vertebral spine may lead to pain and vertebral spine deformations.[99, 100]

TB is in the great majority of cases transmitted by contagion from an infected individual, with active TB, that by coughing, sneezing or exhaling expel aerosols that contain bacilli able to infect another susceptible individual. This type of infection occurs by aerial route, allowing the bacilli to reach the lungs and start a new infection cycle. The infection might still take place vertically during the gestation period by hematogenic path through the umbilical artery, or by the amniotic fluid that is ingested or aspirated in the uterus during birth. This latter type of transmission results in a case congenitus TB that is distinct of acquired neonatal infection that occurs by contact with an individual with respiratory tuberculosis.[101-103]

1.4.2 Infection by *M. tuberculosis*

In the process of infection, *M. tuberculosis* generally gains access to the alveoli through aerosols generated by coughing or sneezing. Larger droplets (> 5 μm) do not reach the alveoli

as they are retained by the ciliated epithelia of the respiratory tract. Only the 1-3 µm sized particles are infectious, can remain airborne for several hours, and are able to reach the alveolar space, each carrying 1-3 bacilli.[104, 105] It is thought that the first contact of the bacillus is with resident macrophages, but it is still thought possible that these bacteria are initially phagocyted by type II pneumocytes.[15] Nevertheless, alveolar macrophages constitute the initial defence against *M. tuberculosis* infection; phagocytosis is mediated by contact of the macrophage mannose receptors and manosyl groups present in the cell wall manosylated LAraM molecules or, through the complement receptors by opsonization of C3 and via CR3.[106, 107]

After phagocytosis, *M. tuberculosis* will initially reside in an endocytic vacuole, designated by phagosome. In normal circumstances, the fusion between pagosomes and lysossomes takes place, generating a hostile environment with an acidic pH, reactive oxygen and nitrogen intermediates (in particular, nitric oxide), lysossomal enzymes and toxic peptides. However, pathogenic mycobacteria inhibit the phagosome-lysossome fusion and still avoid phagossomal acidification.[15, 106, 107] The existing manosylated lipoarabinomannan molecules on the mycobacterial cell wall appear to be important in the acidification blockade since it indirectly interferes, through the phosphatidylinositol 3-cinase, in the accumulation of sintaxin 6 that mediates the incorporation lysossomal hydrolases and vacuolar protonic ATPases in the phagosome.[108, 109] Trehalose 6,6'-dimycolate and the phenolic glycolipids are also essential in blocking the phagossomal acidification.[110]

Infected macrophages, through cytokine production, attract manocytes, lymphocytes and inactive neutrophils that help macrophages in the infection control. $CD4^+$ T lymphocytes recognize processed antigens processed in phagossomes and presented in the form of small peptides by MHC class II molecules in antigen presenting cells such as monocytes, macrophages or dendritic cells. This amplifies the immune response through the production of cytokines that attract more immune cells and, activate effector cells. It is the case for macrophages that once activated, will dispose of several cellular mechanisms to face off the infection such as the reactive oxygen and nitrogen intermediates.[111] $CD8^+$ T lymphocytes on the other hand, recognize peptidic fragments of cytosol-processed antigens exposed by MHC class I molecules, present in all nucleated cells, leading to the lysis or apoptosis of these infected cells.[106] Also important in the host's immune response are the dendritic cells that have the ability to phagocyte *M. tuberculosis* bacilli, and although they don't have a bactericidal capability towards this microorganism, they are able to maintain *M. tuberculosis* in a non-replicative state inside the phagossome. The dendritic cells are antigen presenting cells

that in the lymph node will induce the activation and differentiation of *naïve* T lymphocytes in effector cells. The dendritic cells also produce pro-inflammatory cytokines such as Tumor Necrosis Factor (TNF) and IL-6 that promote the synthesis of TNF and Interferon (IFN)-γ by T lymphocytes, stimulating the macrophage-mediated antimicrobial activity.[110]

It was verified *in vitro* that three different outcomes might occur after phagocytosis: necrosis, in which cell death occurs with plasma membrane degradation; apoptosis, cell death in which the plasma membrane is preserved; or, survival of infected macrophages. Virulent *M. tuberculosis* strains subvert the regulation of prostaglandin E_2 (pro-apoptotic) and lipoxin A_4 (pro-necrotic) eicosanoids, promoting the necrotic pathway over the apoptotic pathway, with the latter being the most important not only in the reduction of bacterial viability but also, in the establishment of T cell-mediated cell immunity.[112, 113] Necrosis will on the other hand facilitate bacilli release and cellular reinfection.[110]

Focal granulomatous lesions are formed, composed by giant cells derived from macrophages and lymphocytes (Figure 1.9). The granuloma is initially composed by an amorphous mass of monocytes, neutrophils and macrophages. The macrophages later differentiate in nucleated giant cells (Langhans Cells), foamy macrophages and, epithelioid macrophages. The appearance of specific lymphocytes only occurs 2-3 weeks after infection. In this way, the propagation of the bacteria is generally contained and as cell immunity develops, bacilli-containing macrophages are killed, leading to the formation of the caseum in the centre of the granuloma, surrounded by a cellular layer composed by fibroblasts, lymphocytes, neutrophiles and monocytes that is correlated with a bacilli-mediated local dysregulation of the host's lipidic metabolism.[115]

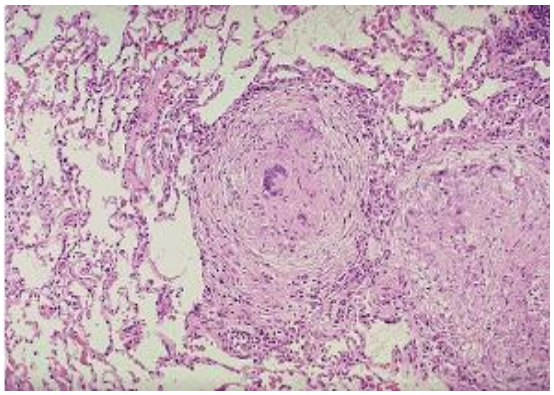


Figure 1.9 – Microphotography of a granuloma and associated necrosis. Extracted from Luna.[114]

The granuloma has traditionally being viewed as a structure designed to protect the host and contain the bacterial replication. Nevertheless, the granuloma might end-up by protecting the bacilli on the inside from the immune system cells' and hamper the action of the antibacillary drugs.[116, 117] In fact, *M. tuberculosis* appears to have an active role in inducing the development of the granuloma by inducing the expression of the host's matrix metaloproteinases, important

in the degradation of the lung cell matrix and, simultaneously inhibit the expression of tissue metalloproteinase inhibitors.[116]

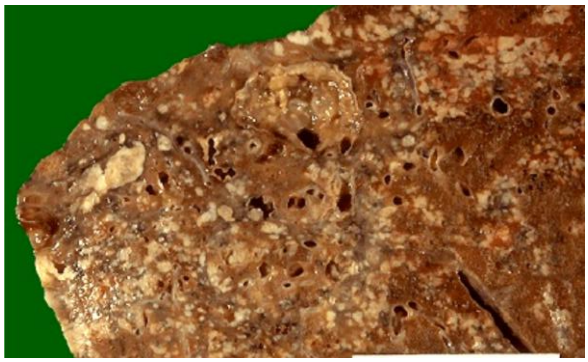


Figure 1.10 – TB affected lung showing necrosis and associated cavitation. Extracted from Luna.[114]

It is thought that *M. tuberculosis* is able to remain in a dormant state for decades on the inside of the granuloma, thus ensuring the persistence of the infection.[118] The foamy macrophages, called this way due to the accumulation of lipidic bodies, loose their bactericidal and phagocytic capability and may act as a niche for bacterial persistence. [117] In fact, it has recently been shown that *M. tuberculosis* induces the foamy phenotype by diverting the glycolytic pathway towards ketone body synthesis, leading to the accumulation of lipidic bodies.[119] The latency state is characterized by a reduced metabolic rate and is triggered by the expression of the DosR regulator, that controls a regulon of approximately 50 genes.[120] The expression of DosR and consequent activation of the regulon is the result of the response to three signals: hypoxia, nitric oxide and carbon monoxide.[121-125]

If the infected individual possesses an efficient cell immunity, the infection can be stopped at this stage, the granuloma heals resulting in small fibrous and calcified lesions.[15] In this granuloma late state, the caseum becomes hypoxic, and might induce a dormancy state or non-replicative persistence.[126] If however, an infected individual with a latent infection develops a weakened immune system due to HIV infection, immunosuppressive drugs, malnutrition, aging, etc., the inside of the granuloma is liquefied by an unknown process and becomes a rich medium for the growth of *M. tuberculosis*. The granuloma rupture contaminates the airways and the propagation of this liquid through the lungs leads to the development of cough, facilitating transmission, turning the individual into an infectious patient (Figure 1.10).[15, 63] It is also possible that bacterial propagation to other tissues may occur by hematogenic or lymphatic way and, cause extrapulmonary TB.[15]

1.4.3 HIV/TB Coinfection

The coexistence of HIV and *M. tuberculosis* in the same host represents at the present one of the more pressing dangers in public health. Each one of these pathogens alters the way in which the other interacts with the immune system potentiating its pathogenicity.

The WHO estimates that in 2010 there were 34 million HIV infected people, having occurred 2.7 million new cases worldwide in the same year. The incidence of the HIV infection has shown a global decreasing tendency since the late 90s but in some regions it appears to have again taken a growing tendency. This is the case of North Africa, Middle East, Eastern Europe and Central Asia.[127]

After the infection by *M. tuberculosis*, the progression rate to active disease is approximately 10% in immunocompetent individuals, half of which develop the disease in the two following years after infection. However, the risk of disease progression in immunosuppressed individuals by HIV infection is 5-15%/year.[128, 129] This substantially higher risk in coinfected individuals results from different HIV pathogenicity inherent mechanisms not only due to the depletion of CD4+ T lymphocytes and consequent decrease of cellular responsiveness. Examples of these mechanisms consist in the reduction of the macrophage oxidative burst, overexpression of *M. tuberculosis* entry cell receptors or the diminishing apoptotic response by macrophages, essential for the elimination of intracellular bacilli.[130, 131]

At the pathological level, while in immunocompetent individuals TB is generally limited to the lungs, in HIV infected patients TB can present as a systemic disease affecting multiple organs. All the forms of extrapulmonary TB have already been described in HIV infected patients. Even at the pulmonary level the majority of the patients do not exhibit the typical granulomatous organization, but more diffuse lesions with granulocytic infiltrate. Cavitary lesions are only rarely found in patients with CD4+ T lymphocitary counts <200/mm^3.[128, 129]

On the other hand, *M. tuberculosis* equally potentiates the HIV pathogenicity inherent mechanisms and disease progression. It has been described that the *M. tuberculosis*-HIV coinfection has as a consequence an increase in the viral replication, which is reflected in a higher viral load in HIV infected individuals with active TB.[132, 133] This increase in viral replication might be due to the induction of TNF and IL-6, caused by the mycobacterial LAraM, that through the NF-κB pathway leads to the transcriptional activation of the proviral long terminal repeats (LTR) promoter.[129]

M. tuberculosis also promotes the HIV infection through the overexpression of the CXCR4 and CCR5 coreceptors; induction of pro-inflammatory cytokine release, mainly the TNF; and, the decrease of the CCL5 expression.[130] Ku *et al* also demonstrated that TB affects the $CD4^+$ T lymphocyte count kinetics since in co-infected patients, the recovery of $CD4^+$ T lymphocyte to levels >250 cells/mm^3 is delayed.[134]

1.5 TUBERCULOSIS DIAGNOSIS

TB diagnostic is a process involving several stages and can be divided in two phases: the initial and definitive diagnostic. The initial diagnostic is done based on clinical data such as patient's symptoms and radiological data.

Definitive diagnostic implies a laboratory diagnosis performed by mycobacteriology specialized laboratories using the patient's clinical samples, *e.g.*, sputum or biopsy. The present criteria for laboratory confirmed TB cases imply a positive culture for MTC or, positive microscopy for acid-fast bacilli and detection of MTC specific DNA through a nucleic acid amplification test (NAAT).

1.5.1 Microscopy

Microscopy is based on the differential detection of acid-fast bacilli. This property is not exclusive of mycobacteria and it might be further observed in member species from the *Actinomyces*, *Nocardia*, *Corynebacterium*, *Rhodococcus* genera and, in oocysts from *Cryptosporydium* spp, *Isospora belli* and *Carcystis* spp.

The development of the acid-fastness staining method applied to TB diagnostics had its beginnings in the late 19th century with the works of Robert Koch, Georg Rindfleisch, Paul Ehrlich and, subsequent development by Franz Ziehl and Friedrich Neelsen for acid-fast bacilli (AFB) detection.[135] The Ziehl-Neelsen staining consists in a hot carbol-fucsin stain, decolorizing with acid and conterstaining with methylene blue, malachite green or picric acid. Several variations of this Ziehl-Neelsen stain have come up through the years, usually differing on the decolorizing agent and solvent. The acids usually employed for the decolorizing purpose are the sulphuric acid at 25%, diluted in water, or chloridric acid at 3% in ethanol.[99, 136]

Three major modified versions of the Ziehl-Neelsen staining procedure exist and are the most widely used: the traditional Ziehl-Neelsen staining, the Kinyoun staining and, Truant's Auramine-Rhodamine staining. The Kinyoun staining is a modified version of the Ziehl-Neelsen staining, very similar to the original, except the fact that the carbol-fucsin solution is used at a much higher concentration avoiding the use of heating in the primary staining procedure.[99, 136] However, in a study by Somoskövi *et al* comparing the three techniques has shown that the Kinyoun staining was the only that has produced false-negative results, while the Ziehl-Neelsen and the Auramine-.Rhodamine stainings have shown comparable specificities and sensitivities.[137]

The Auramine-Rhodamine stain uses two fluorescent dyes (Auramine O and Rhodamine), although Auramine O can be used alone, by which a fluorescence microscopy is required. Slide examination time is reduced since it can be performed with lower magnifications (250X *vs* 1000X), and the same preparation can be confirmed and examined using the Ziehl-Neelsen stain. Ba *et al* has demonstrated that the mean time to declare a slide as negative using the Auramine-Rhodamine stain was 3 minutes and 44 seconds *vs* 7 minutes and 44 seconds using the Ziehl-Neelsen staining procedure.[138]

Microscopy has, nevertheless, a reduced sensitivity when compared with culture detection, only allowing the detection in samples with 10^4-10^5 AFB/ml.[139] In the microscopy context, the concentration by centrifugation assumes special importance in the technique's sensitivity.[140]

1.5.2 Culture

Culture isolation of clinical isolates or strains of any MTC member is a time-consuming process that requires specially developed culture media. The TB bacillus was initially isolated in coagulated serum. Presently, diverse culture media are employed for this purpose, synthetic or complex and selective or non-selective.[99, 136]

Comparing with microscopy, culture isolation allows to posteriorly identify the isolate, eliminating the non-specificity problem of microscopy and, shows a higher sensitivity as it allows detection in samples with 10-100 AFB/ml.[139]

There are several formulations for the egg-based complex media, the most commonly used being the Löwenstein-Jensen (LJ), with several variations, *e.g.*, the Ogawa medium (no

asparagine) or Coletsos. This type of media are solidified by inspissation at 80-85ºC and include, similarly to agarized media, malachite green, that confers some selectivity to the medium avoiding the growth of contaminating bacteria. The Petragnami medium, mostly destined to the isolation of mycobacteria from samples with a higher contamination degree, has a higher malachite green concentration. The American Trudeau Society medium (ATS Medium) has, on the other hand, a decreased concentration of malachite green and is therefore mostly destined to the isolation of strains from usually sterile clinical samples, *e.g.*, cerebrospinal fluid.[136]

Posteriorly to LJ, several formulations of synthetic media have been developed, mostly by the works of Middlebrook, Dubos and Cohn.[141] The Middlebrook 7H9 (liquid) and 7H10 (solid, agarized) are the most diffused culture media, having the advantage, regarding the previous media, to allow the incorporation of antibacillary drugs for drug susceptibility testing. These media are usually supplemented with glycerol (2%, except for *M. bovis*), oleic acid, albumin, dextrose and catalase. There is also the 7H11 medium that differs from 7H10 by containing 0.1% casein hydrolysate, allowing an improved recovery of INH-resistant isolates. Both types of media can be supplemented with antimicrobial agents to ensure a higher selectivity, *e.g.*, the LJ Gruft contains nalidixic acid and penicillin.[136]

Growth on solid medium is however a time-consuming process due to the low replication rate of all members of the MTC. In the last decades, (semi-)automatized liquid culture systems have been developed, allowing a faster detection and diagnosis.

The two most widely used systems have been the BACTEC™ 460 (Becton Dickinson™) and the BACTEC™ MGIT™ (Mycobacteria Growth Indicator Tube) 960. The first is based on the radiometric detection of $^{14}CO_2$ released due to the metabolism of palmitic acid radioactively labelled. The BACTEC™ 460 is however at a discontinuation stage but has been the gold standard for more than two decades.

The BACTEC™ MGIT™ 960 or the 320 version, more recent and with a lesser capacity, is now considered the gold standard for mycobacteria culture detection. This system is based on the detection of emitted fluorescence by a compound present in the culture tube that is O_2 sensitive. As the O_2 is consumed, the fluorescence increases indicating bacterial growth.

The mean time to detection in both systems is comparable (10-15 days) but is significantly less than growth detection on solid medium (20-26 days), particularly on AFB negative smears (16-18 *vs* 42 days).[142-144]

1.5.3 Drug Susceptibility Testing

Several methods have been traditionally employed in testing drug susceptibility to the different antibacillary drugs used in TB treatment, namely, the absolute concentration method, the resistance ratio method and the proportion method.[145] The proportion method has been adopted as the gold standard for TB drug susceptibility testing (DST). The rationale behind the proportion method is to assume that regarding resistance/susceptibility phenotype, natural variability exists in a *M. tuberculosis* population in its host but, if not more than a critical proportion (usually 1%, depends on the media and drug tested) of the population is able to grow in the presence of a drug, the calculated proportion is not of clinical significance and the strain is considered susceptible to the action of the tested drug.[145]

Standard DST by the proportion method was originally performed on egg-based media and, later, on agar-based media (preferred, *e.g.*, Middlebrook 7H10 or 7H11) through inoculation of several inocula dilutions in drug-containing and drug-free media and count of colony forming units (CFUs).[145, 146] The proportion of resistant CFUs is calculated by dividing the number of CFUs in drug containing medium by the number of CFUs in the drug free control, multiplied by 100. If the resistant CFUs account for more than the considered critical proportion of the total CFUs that develop on drug-free media, the strain is considered susceptible to that drug.[145-147] The concentration at which the drug is tested is called critical concentration, and is used as a breakpoint for classification of a strain as resistant or susceptible. This concentration is defined as the concentration that inhibits ≥95% of wild-type strains not previously exposed to the drug.[148]

The proportion method has also been adapted for commercial liquid-culture systems, such as the BACTEC 460 and BACTEC 960, in which the bacterial suspension is inoculated in a vial or tube with drug containing medium at the appropriate critical concentration. Furthermore a growth control is also established through the inoculation of a 1:100 dilution of the bacterial suspension. Briefly, if the bacterial growth rate in the drug medium is higher than the growth rate registered for the control vial or tube, the strain is considered resistant.[146] Moreover, the liquid proportion method is less likely to be as influenced by bacterial clumping as the proportion method in solid media.[149]

Susceptibility testing to PZA is performed differently than for other drugs as PZA is only active at an acidic pH and most reliable results are obtained using commercial systems with slightly acidic medium (pH=6).[146] The least reliable PZA drug susceptibility test is the one performed on acidic 7H10/11 supplemented with albumin, dextrose and catalase (no oleic acid as it is toxic at an acidic pH).[150]

A quantitative DST assay has been more recently developed through the use of BACTEC 960 equipped with specially developed software that monitors and automatically stores the growth units and plots the growth rates.[151] Comparison of growth rates in presence of different drug concentrations against a 1:100 diluted growth control allows classification as resistant, susceptible or intermediate resistant for each of the tested drug concentrations and, assess the strains' level of resistance (low-, intermediate- or high-level).[151]

1.5.4 Nucleic Acid Amplification Tests

Due to the fact that the detection of the TB bacillus in clinical samples is a time consuming process and bacilloscopy is a method with a reduced specificity, nucleic acid amplification tests (NAAT) are equally employed in laboratory diagnosis. The NAAT are based on PCR amplification and detection of specific DNA or RNA from the MTC members.

Early NAAT tests included the Amplicor™ (Roche™) and the Enhanced Mycobacterium tuberculosis direct (E-MTD™) test (Gen-Probe™). The Amplicor test was based on the amplification and MTC-specific detection of the 16S rRNA gene, but has been discontinued. The E-MTD test is now on its second version, the amplified MTD test, and is based on the extraction, amplification and MTC-specific detection of rRNA by chemiluminescence. For this latter test, several studies report overall sensitivities of about 89-100%, although decreased susceptibilities are reported on negative AFB samples.[152-162]

Presently, the more diffused tests are based on specific amplification and reverse hybridization in strips with immobilized probes, *e.g.*, GenoType MTBSRplus (Hain Lifescience) and INNO-Lipa Rif. TB (Innogenetics™). Both tests allow detection from the clinical sample and screen for mutations associated to RIF and, for GenoType MTBDRplus, also with INH. Several published studies appear do demonstrate that the sensitivity of both methods is comparable (82-95%), although usually inferior in negative AFB samples or non-respiratory samples with decreased bacillary load.[163-171] Also commercialized are tests that allow the sole detection of MTC species

(GenoQuick MTB, Hain Lifescience) or specific non-tuberculous mycobacteria (GenoType Mycobacteria Direct, Hain Lifescience).

In recent years a new automatic system, GeneXpert™ MTB/RIF (Cepheid™), in which amplification and detection is performed in an almost fully automated way in a cartridge through a microfluidics system based on real-time PCR.[172] This system allows detection of MTC bacilli and RIF associated mutations in two hours, with a minimal hands-on time. Several studies have demonstrated specificities of approximately 99% and sensitivities above 90% on AFB positive samples, but decreased sensitivities in extrapulmonary or AFB-negative samples.[173-188]

In a recent study by Skenders *et al* it was demonstrated that in a total of 107 patients with positive bacilloscopy, the implementation of the INNO-Lipa Rif. TB test in the diagnosis routine has allowed to speed-up the empirical treatment but not the time or rate of culture conversion.[189]

1.5.5 Immunological Testing

Immunological testing presents an alternate/additional route to TB diagnostics, particularly in the case of latent tuberculosis infection (LTBI). The traditional method for LTBI diagnosis has been the tuberculin skin test (TST), in which an *M. tuberculosis* purified protein derivative (PPD) or tuberculin is usually injected intradermically. If the individual has come in contact with the tubercle bacillus an induration is expected to develop as a result of antigen recognition by the immune system. The early tuberculin or old tuberculin was initially developed by Robert Koch and consisted in a protein mixture extracted from boiled *M. tuberculosis* cultures. As this old tuberculin contained variable proportions of different proteins, which depended on the production method, a standardized tuberculin preparation was developed by Seibert *et al*, designated Purified Protein Derivative (PPD) and the international standard designated as PPD-S (standard) or PPD-M (mammal).[190] TST has however known limitations regarding cross-reactivity with other environmental mycobacteria and BCG vaccination.

In the last decade, the development of IFN-γ release assays (IGRA) has resulted in two commercial tests for TB and LTBI diagnosis: QuantiFERON™-TB Gold In-tube (QFT-GIT) (Cellestis Ltd, Carnegie, Australia) and T-SPOT™.TB (T-SPOT) (Oxford Immunotec, Abingdon,

UK). QFT-GIT is a whole-blood test based on Enzyme-Linked Immunosorbent Assay (ELISA) to quantitatively detect IFN-γ produced by lymphocytes upon overnight exposure to three TB antigens: ESAT-6, CFP-10 and TB7.7. In T-SPOT, peripheral blood mononuclear cells (PBMCs) are separated, counted and exposed to two TB antigens: ESAT-6 and CFP-10. The production of IFN-γ is assessed by culture supernatant ELISA in QFT-GIT and by enzyme-linked immunospot (ELISpot) in T-SPOT. In both IGRA tests no cross-reactivity with BCG vaccination is expected since all antigens that are used in both tests belong to the RD1 (ESAT-6 and CFP-10) or RD11 (TB7.7) regions that are absent from *M. bovis* BCG strain.

Recent systematic reviews and meta-analysis of published data show that overall pooled sensitivity may range between 61-69% for QFT-GIT and 65-66% for T-SPOT although stratification by high/low-burden setting may yield slightly different results: in high-burden settings, 61-65% for QFT-GIT and, 65-72% for T-SPOT; in low-burden settings, 59-67% for QFT-GIT and, 69-94% for T-SPOT.[191-194] Overall specificity ranged between 72-76% and 63-70% for QFT-GIT and T-SPOT, respectively. In low-burden settings, specificity ranged between 89-94% for QFT-GIT and was 64% for T-SPOT, while in high-burden settings it ranged between 50-62% and 52-73% for QFT-GIT and T-SPOT, respectively.[191-194] Except for Chen *et al* that found statistically significant higher specificity value for QFT-GIT in low-burden settings, none of the differences reported in sensitivity or specificity is statistically significant.[191-194]

Although T-SPOT is generally regarded as a more sensitive technique, the exclusion of failed tests due to insufficient number of PBMCs may contribute to the overestimation of the test's sensitivity in HIV-infected patients.[191]

Another important aspect is the proportion of indeterminate results and several factors can affect this parameter: younger age; malaria, HIV or helminth infection; iron deficiency; iron deficiency anaemia or HLA allele.[195, 196] In the same recent systematic reviews and meta-analysis, overall pooled proportion of indeterminate results was 8.2-10.0% and 5.9-13.2% for QFT-GIT and T-SPOT, respectively. Although both tests had generally higher proportion of indeterminate results in high-burden settings: 4-15% (QFT-GIT) and 2-14% (T-SPOT) *vs* 4-8.4% (QFT-GIT) and 0-5% (T-SPOT), for high- and low-burden settings, respectively, no significant differences were found.[191-194] Santin *et al* found a statistically significant higher proportion of indeterminate cases using QFT-GIT among patients with active TB (culture confirmed or symptomatic), 15.3%, in comparison with patients screened for LTBI, 3.9%.[191] A higher, but not statistically significant, difference in the proportion of indeterminate results was also found for

QFT-GIT (8.1%) in comparison with T-SPOT (4.0%) when using a patient stratification threshold of 200 CD4+ cells/ml.[191]

Most studies compare the performance of IGRA tests in patients with active TB and the data regarding the performance of these tests in diagnosing and screening LTBI is limited due to the lack of a gold-standard. However, two studies by Aichelburg *et al* and Santin *et al* showed that QFT-GIT had a 100% negative predictive value for development of active TB.[197, 198]

1.6 TUBERCULOSIS TREATMENT

1.6.1 Tuberculosis Treatment Regimens

The antibacillary drugs available to fight TB are divided into first-line drugs and second-line drugs. The former present a higher efficacy, less side effects and are comprised by INH, RIF, ethambutol (EMB), STP and PZA. Second-line drugs are not as efficacious as first-line drugs, present a higher degree of side effects and include several drugs such as: KAN, AMK, CAP, FQs, ethionamide (ETH), para-amino salicylic acid (PAS), cycloserine (CS), clofazimine (CFZ) and prothionamide (PTH).

For TB drug therapy design it is necessary to acknowledge that different bacilli subpopulations exist in the host, concerning the bacterial metabolic state. According to the Mitchison's hypothesis, there are four different populations of bacilli with distinct metabolic states: the first, exhibit continuous growth and are actively dividing, comprising the large majority of the bacilli within the host with active TB; the second, is comprised by bacilli in an acidic environment, within macrophages or the caseum, in a state of semi-dormancy; the third subpopulation is composed by bacilli located extracellularly in solid caseous lesions, that exhibit intermittent growth; the fourth subpopulation is comprised by bacilli in a complete dormant state (persisters) that is, at the present, beyond the reach of any drug.[199, 200] To tackle these populations some of the available antibacillary drugs are more appropriate than others (Figure 1.11).[199, 200]

TB chemotherapeutic regimens can be divided in two stages or phases: the initial phase, that includes a higher number of drugs to achieve a rapid killing effect and sputum conversion, with the main purpose of reducing the patient's infectiousness; and, a longer continuation phase, with less drugs, that aim at a better sterilization effect and prevention of relapses.[200]

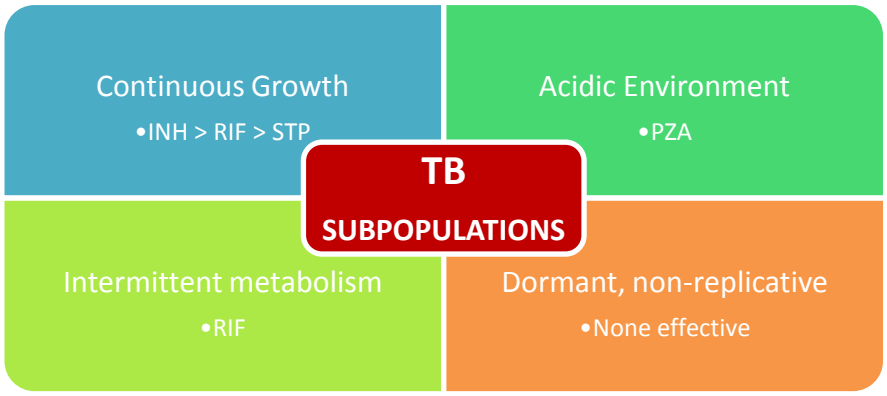


Figure 1.11 - Schematic representation of the four different bacilli subpopulations that exist in a host and the first-line drugs that better target these, according with the Mitchison's hypothesis. Adapted from Khadivaran & Sharma, 2008.[201]

The cornerstone of TB treatment is the standard six-month regimen constituted by a two-month initial phase of INH+RIF+PZA+EMB and a four-month continuation phase of INH+RIF (former Category I). This regimen is recommended for every new patient with no MDR-TB suspicious, although EMB should be maintained throughout the continuation phase in settings with high primary drug resistance rates (*i.e.*, drug resistance not resulting from a previous treatment).[202, 203]

Another first-line treatment regimen, formerly Category II, was recommended in which STP or EMB were added alone, and both treatment phases extended by one additional month for relapses or re-treatments due to patient default. In the most recent recommendations, category II regimen with the addition of STP alone is still recommended, albeit with a different designation, for previously treated patients, including relapses and defaults, where the likelihood of MDR-TB is low or medium. If DST data becomes available this regimen should be adapted or modified; if not, it is recommended by the WHO to maintain the regimen for the full course of the treatment. Whenever the likelihood of MDR is high on a previously treated patient, it is recommended to go for an empirical MDR-TB treatment regimen.[202, 203]

The previously standard regimens recommended by the WHO for MDR-TB treatment of patients with suspicious MDR-TB, but with no DST data available or DST data for first-line drugs only are schematized in Table 1.5.[202]

Table 1.5 – MDR-TB treatment regimens recommended by the WHO. Adapted from WHO guidelines.[203]

First-line DST	Initial Phase		Continuation Phase	
	Drugs	Duration (months)	Drugs	Duration (months)
Not available	KAN+ETH+FQ+PZA+/-EMB	>=6	ETH+FQ+PZA+/-EMB	12-18
INH+RIF	STP+ETH+FQ+PZA+/-EMB	>=6	ETH+FQ+PZA+/-EMB	12-18
All first-line drugs	1 Injectable Drug+ 1FQ + 2 of: PAS, ETH, CS	>=6	All except the injectable drug	18
Second-line drug DST available	Tailored regimen according with DST pattern			

According with the most recent WHO recommendations, each country should choose the most adequate MDR standard treatment regimen, based on surveillance data and drug resistance surveys.[203]

Table 1.6 – Hierarchy of five drug groups that should be chosen for tailored MDR-TB treatment regimens. Extracted from WHO guidelines.[203]

Drug Groups				
1	**2**	**3**	**4**	**5**
PZA	KAN	Levofloxacin	PAS	CFZ
EMB	AMK	Moxifloxacin	CS	Linezolide
Rifambutin	CAP	Ofloxacin	Terizidone	Amoxicilin/Clavulanate
	STP		ETH	Thioacetazone
			PTH	Imipenem/Cilastatin
				High-dose INH
				Clarithromycin

For tailored regimens the WHO has defined five groups of drugs, that should be added to a regimen in a hierarchical order (Table 1.6).[203] At least four effective drugs should be added to a regimen following the group order. Only one drug can be chosen from groups 2 and 3, while as many drugs a possible, or necessary, can be chosen from the remaining groups. The remaining general principles for MDR-TB treatment design can be found further detailed in the most recent WHO guidelines.[203]

1.6.2 Antibacillary drugs

INH

INH or isonicotinic acid hydrazide is a synthetic prodrug that requires activation by the bacterial catalase peroxidase, encoded by the *katG* gene.[204] Upon activation an isonicotinic acyl radical is produced that binds NAD, forming an INH-NAD adduct that targets the InhA enoyl-ACP reductase, essential for mycolic acid elongation by the type II fatty acid synthase system (FASII).[205-207] The inhibition of mycolic acid biosynthesis leads to cell lysis and death. INH enters the cell by passive diffusion and its activity is restricted to dividing bacteria under aerobic conditions.[208]

INH great efficacy is in part due to its low minimum inhibitory concentration (MIC): 0.02 mg/L for *M. tuberculosis* H37Rv and 0.02-0.05 mg/L in susceptible clinical isolates.[209] INH has a bactericidal activity against rapidly growing mycobacteria and bacteriostatic against slow-growers, although bactericidal activity is observed against *M. tuberculosis*, *M. bovis* and *M. kansasii.*

INH resistance usually develops as a result of *katG* mutations that decrease the ability of the catalase-peroxidase to convert INH to its active form. Depending on the mutations found, *katG* mutations may result in low to high-level resistance.[210] Resistance level is negatively correlated with the remaining KatG INH oxidase activity.[210] The most common mutation found in *katG* is a serine to threonine substitution at codon 315 (S315T), found in up to 93% of INH resistant isolates, associated with an INH high-level resistance.[204, 211-234] S315T has been shown to decrease KatG catalytic activity and binding affinity towards INH.[235-237]

Another important mechanism of INH resistance, usually second to *katG* mutations, consists in the acquisition of hypermorphic mutations in the promoter region of *mabA(fabG1)-inhA* operon in up to 32% of INH resistant isolates.[215-221, 223, 224, 226-232, 234, 238] Although such mutations usually lead to INH low-level resistance, an unusual high-prevalence of *inhA* promoter mutations have been detected in Lisbon, Portugal, and in *M. africanum* West-Africa 1 lineage.[24, 223, 225]

Furthermore, NAD^+/NADH altered ratios, mediated by *ndh* mutations, can also result in INH and ETH resistance.[207] Mutations found in other genes, namely *kasA* and *ahpC*, have an unknown or questionable role in INH resistance.[223, 227, 230, 232, 234] Some studies have however

reported up to 35% of INH resistant clinical isolates without mutations in any of the above genes, suggesting alternate mechanisms for INH resistance, such as drug efflux.[227, 239-241]

RIF

RIF or rifampin is a semi-synthetic antibacillary drug derived from rifamycin introduced in TB chemotherapy in 1967.[242] RIF binds to the β-subunit of the DNA-dependent RNA polymerase, encoded by the *rpoB* gene.[243] Binding to the RNA polymerase β-subunit is thought to physically block transcription of growing RNA chains when these become 2-3 nucleotides in length.[244, 245] RIF has a bactericidal activity against most Gram-positive, some Gram-negative and mycobacteria.[246, 247] Active against metabolically active bacteria, RIF also possess some degree of sterilizing activity as it is active against latent bacilli with spurts of activity.[242, 248]

RIF MIC ranges between 0.2-0.4 mg/L for susceptible clinical isolates (0.4 mg/L for *M. tuberculosis* H37Rv).[209]

Acquisition of RIF resistance is usually the result of aminoacid substitutions in a 81-bp region of the rpoB gene named RIF resistance determining region (RRDR) or cluster I.[249-251] Besides aminoacid substitutions, deletions or insertions in *rpoB* have been reported in some studies.[252, 253] The most common substitutions occur in codons 531 (prevalence of 31.0-76.9% of the RIF-resistant isolates), 526 (7.7-43.0%) and 516 (3.4-28.6%).[214, 215, 223, 230, 231, 245, 249, 250, 252-261] Mutations in these three codons are thought to have the least impact on fitness, particularly S531L mutations.[262, 263]

RIF resistance level depends on the substituted residue, or combination of mutations, and the type of substitution, *e.g.*, S531L, H526D or D516V result in high-level resistance whereas D516Y results in low-level resistance.[264-266]

Up to 11% of RIF resistant isolates do not show any RRDR mutations, with some isolates displaying N-terminal mutations that may affect protein-drug interaction or, mutations in *rpoB* cluster II (codons 571-572).[249, 261]

A different mechanism involving inactivation of RIF by ribosylation in mycobacteria has been identified as partially responsible for the low susceptibility of some species to RIF, but no mechanism of this type has been described for *M. tuberculosis*.[267, 268]

EMB

EMB is an antimycobacterial drug synthesized from ethylenediamine and used in TB treatment since 1968.[242] EMB targets the cell wall biosynthesis, more specifically inhibiting the arabinosylation of cell wall arabinogalactan (AG) and LAraM.[269] EMB targets the arabinosyltransferases encoded by the *embCAB* operon, but inhibiting the activity of EmbB with a greater affinity than EmbC. Since e*mbAB* gene products are responsible for the arabinosylation of AG whereas arabinosylation of LAraM is catalyzed by the *embC* gene product, EMB leads to a more rapid inhibition of AG biosynthesis than LAraM biosynthesis.[270-273] The EMB mediated AG biosynthesis inhibition leads to further mycolic acid accumulation due to the depletion of mycolate attachment sites.[270]

EMB enters the cell in a passive manner and, has a bacteriostatic activity against metabolically active bacilli rather than bactericidal.[274-276] EMB MICs range between 0.5-2 mg/L for susceptible isolates (0.5 mg/L for *M. tuberculosis* H37Rv).[209]

EMB resistance has been traditionally associated with *embB* mutations, of special incidence on a EmbB 142 aa extracytoplasmatic loop.[277-280] The most common mutations occur in codon 306 in up to 68% of EMB resistant isolates and usually involve the substitution of a methionine for a valine, leucine or isoleucine.[219, 281-290] Strains bearing *embB*306 mutations have been associated with a higher level of EMB resistance than other *embB* mutations.[286, 290, 291] On the other hand, M306V/L mutations appear to confer a higher level of EMB resistance than M306I.[219, 278, 285]

In more recent years the utility of *embB*306 mutations as molecular determinants of EMB resistance has been questioned as several authors reported EMB susceptible isolates bearing *embB*306 mutations resistance whereas others only detected *embB*306 mutations among EMB resistant isolates.[219, 283, 284, 287, 292-294] Allelic exchange experiments have shown that *embB*306 mutations only moderately increase EMB MIC to levels below the ones observed in clinical isolates. Moreover, *embB*306 mutations appear to be required for high-level resistance, highlighting that this level of resistance is the result of multigenic mutational events.[295-297] Other genes such as the *iniA* gene, essential to EMB efflux, may be involved in EMB resistance although the paucity of data in this regard prevents such associations.[278, 298, 299]

STP

STP was the first antibiotic used in TB treatment, in 1944, and is an aminoglycoside isolated from *Streptomyces griseus*.[242, 300, 301] STP interferes with protein synthesis by inhibiting genetic translation with a more pronounced effect on elongation rather than initiation.[211] Crystallography of *Thermus thermophilus* 30S ribosomal subunit complexed with STP, shows that STP tightly binds 16S rRNA (*rrs* gene) through both salt bridges and hydrogen bonds to four different parts of the molecule and across ribosomal protein S12 (*rpsL* gene)[302] The binding of STP appears to stabilize the ribosomal ambiguity state (ram state) and have a negative effect on the ribosomal proof-reading activity.[302, 303] STP-induced mistranslation leads to an increased proteolysis rate and abnormal misread proteins appear to create protein channels causing loss of permeability control, accelerating STP irreversible uptake and block of translation initiation.[304-307] The stabilization of the peptidyl-tRNA in the A-site has also been demonstrated to reduce the translocation rate.[308]

The initial use of STP in monotherapeutic regimens has lead to the emergence of high number of resistant cases, and with subsequent development of more efficacious antibacillary drugs, STP gradually lost its role as a first-line drug, mainly in developed countries.[309]

STP MICs range between 1.0-2.0 mg/L (1.0 mg/L for *M. tuberculosis* H37Rv) and has a moderate bactericidal activity against susceptible isolates.[209, 301] On comparison with the two other aminoglycosides used in TB treatment, KAN and AMK, STP is the least toxic.[310]

Resistance towards STP is usually mediated by mutations in the *rpsL* gene, generally a substitution of a lysine by an arginine at codon 43, although mutations on codon 88 are often reported and associated with STP resistance.[311] Mutations in *rpsL* gene are associated with a high-level resistance, particularly K43R mutations.[312-315] Another mechanism of STP resistance is *rrs* gene mutations that generally occur in the 530 and 910 loops, which interact with STP and are in close proximity due to the 16S rRNA secondary structure.[311] Mutations in the *rrs* gene usually yield a resistance level below the resistance level of isolates bearing *rpsL* mutations, but high enough to be considered an intermediate resistance level, although MICs can vary significantly depending on the mutation.[313, 314]

The prevalence of each of these mutations displays a wide variability, depending on strain lineage or geography. Mutations in *rpsL* gene are usually more prevalent (up to 89%) than *rrs* mutations (0 – 28.6%).[220, 287, 312, 315-324] Nevertheless, in a study by Cuevas-Cordoba *et al, rrs*

mutations were found to be more prevalent than *rpsL* mutations, 28.6% *vs* 19.8%, respectively.[317]

The loss of the *gidB*-encoded rRNA methyltransferase function, found to be responsible for the methylation of G527 of the 16S rRNA, leads to a reduced affinity of STP towards the ribosome and STP low-level resistance.[325, 326] GidB mutations have been detected in clinical isolates, although the precise quantitative role in STP resistance is still unclear as some of the mutations detected in STP-resistant isolates are also present in susceptible ones.[214, 327] Other polymorphisms detected in *gidB* have been associated with strain lineage rather than STP resistance.[327, 328]

PZA

PZA is a nicotinamide synthetic prodrug that has been introduced in TB treatment in 1970 and is important in the short-course TB treatment.[242] PZA is thought to enter the cell by passive diffusion where it is converted by the bacterial pyrazinamidase (PZase)/nicotinamidase into pyrazinoic acid (POA), that accumulates in the cell due to a deficient efflux system, only leaving the cell by passive diffusion in an acidic pH environment. Extracellular protonation of POA allows reentrance by passive diffusion where deprotonation leads to cytoplasm acidification.[150, 329, 330] The acidification of the cytoplasm may contribute to the depletion of the membrane energy potential.[331]

For many years the identification of a precise target of PZA was unsuccessful as attempts to isolate POA resistant mutants have failed.[332, 333] More recently, Shi *et al* found that PZA inhibits trans-translation through the targeting of the *rpsA*-encoded ribosomal protein S1 by POA.[334]

PZA has an excellent sterilizing activity as it is bactericidal against semi-dormant bacilli in an acidic pH. The introduction of PZA in conjunction with RIF has showed synergism and allowed the shortening of treatment regimens containing both drugs from 12 to 6 months.[150, 335]

The main mechanism behind PZA resistance is the acquisition of mutations in the *pncA* gene.[336] Mutations in the *pncA* gene have been identified in numerous studies in about 72.0-99.9% of the PZA resistant isolates studied.[214, 219, 332, 337-354] Most of the mutations identified are predicted to affect the enzymatic activity of PncA due to the alteration of the active site or destabilization of the protein structure leading to the loss of the PZase activity.[353] As the *pncA* gene is not essential to bacterial viability, the gene is under a less stringent mutational constraint leading to a high mutational diversity and independent arise of PZA resistance at a

lower rate (10^{-5}).[340, 342, 353, 355] The resistance level to PZA depends on the mutations, although mutational hotspots have been identified, with the higher proportion in the loop between the β2 and β3 strands that comprise the metal coordination site of the enzyme's catalytic centre.[332, 346, 356] In fact, docking simulations show that some point mutations contribute to a higher resistance level than others.[357]

PZA resistant isolates with a wild-type *pncA* gene fall in two groups: PZase-negative and PZAase-positive isolates.[338, 339, 347] While PZA resistance in PZase-positive isolates may be explained by mutations at other *loci*, namely the target of POA, resistance in PZase-negative isolates with a wild-type *pncA* gene is probably due to a differential regulation of *pncA* gene expression.

Mutations in RpsA may also lead to PZA resistance through the inhibition of POA-RpsA binding.[334] Nevertheless, Alexander *et al* suggested that *rpsA* mutations are not robust markers for PZA resistance.[337]

Injectable Second-Line Drugs

The injectable drugs of second-line drug use in TB treatment are KAN, AMK and CAP. Although KAN and AMK are, such as STP, aminoglycosides and CAP is a macrocyclic peptide from the tuberactinomycin family, these drugs share the same basic action mechanism and incomplete cross-resistance between the three has been well documented.[358, 359] KAN, AMK and CAP have shown *in vitro* bactericidal activities. Furthermore, CAP has a bactericidal effect against non-replicating *M. tuberculosis*.[360]

These three drugs act through the inhibition of the genetic translation due to ribosomal binding. Nevertheless, the exact mechanism through which these drugs inhibit the ribosomal activity during translation is not fully understood or studied to the extent that has been with STP. AMK and KAN have been shown to bind to the A-site of the ribosome and inhibit translation.[361] CAP mode of action has been initially deduced from early studies of ribosomal translation inhibition with its analogue viomycin (VIO) and, more recently confirmed by the crystal structure of CAP and VIO in complex with the 70S ribosomes.[362-364] These studies have shown that both VIO and CAP affect the dissociation of the 70S ribosome of *M. smegmatis* by stabilization of the 70S couples and, inhibit the translocation by arresting the peptidyl-tRNA in the ribosomal A-site.[362-364] CAP binds to the interface between helix 44 and helix 69 of the small (16S) and large (23S) ribosomal subunits, which is dependent on the methylation, by the

TlyA 2'-O-methyltransferase, of C1409 in helix 44 and C1920 in helix 69 of the 16S and 23S rRNA, respectively.[363, 365]

Mutations in the 16S rRNA-encoding *rrs* gene can mediate cross-resistance between CAP, KAN and AMK. The most common mutation found in the *rrs* gene is the A1401G mutation, present in about 49.3-88.6% of the resistant isolates, and associated with resistance to KAN and AMK, and CAP low-level resistance.[215, 219, 234, 359, 366-370] The C1402T mutation does, on the other hand, mediate resistance to KAN, CAP, VIO but not AMK whereas the G1484T mutation mediates resistance to all four drugs (KAN, AMK, CAP and VIO).[359]

KAN resistance has also been linked with *eis* overexpression, a gene encoding an acetyl transferase that can multi acetylate several aminoglycosides, rendering the drugs ineffective.[371] Since Eis acetylates KAN more efficiently than AMK, *eis* overexpression has been associated with KAN low-level resistance but not AMK resistance.[372] In this regard, several *eis* promoter mutations have been identified in clinical isolates.[373, 374]

CAP resistance can also be mediated by *tlyA* mutations.[373, 375] Maus *et al* have shown that loss of TlyA 2'-o-methyltransferase function causes CAP resistance and, Manshupanee *et al* have shown that TlyA N-terminal aminoacid substitutions reduce its ability to methylate C1409 and C1920 in the 16S and 23S rRNA and that subtle changes in the level of rRNA methylation lead to significant differences in the CAP sub-inhibitory concentrations.[375, 376] Nevertheless *tlyA* mutations in CAP resistant isolates are only seldom observed in CAP resistant isolates as it is more likely that CAP resistance develops as a result of KAN/AMK cross-resistance.[368, 373] Another possibility is that CAP resistance may be mediated by 23S rRNA mutations.[365]

FQs

FQs are fluorinated quinolones at the central ring system. These are broad-spectrum antibacterial drugs that target the bacterial DNA gyrase and topoisomerase IV, therefore inhibiting DNA replication.[377] *M. tuberculosis* lack the toposisomerase IV *parE* and *parC* subunit homologues, limiting the inhibitory activity of FQs to the DNA gyrase through interaction with both GyrA and GyrB subunits.[62, 378] Quinolones bind to the DNA Gyrase-DNA binary complex producing a ternary complex that halts DNA replication leading to cell death.[379] The binding site of quinolones - the Quinolone Binding Pocket (QBP), constituted by protein and DNA, is located in the enzyme's catalytic core where the drug interacts with both the GyrA and GyrB subunits and is intercalated between the dinucleotide step.[379]

FQ resistance has been reported to be increasing in several settings and several studies have described an association between previous exposure to FQ prior to TB diagnosis and an increased risk of having FQ-resistant TB.[380, 381] Devasia *et al* also associated a previous FQ treatment for more than 10 days with FQ resistance.[382] This data seem to correlate with the fact that quinolones induce the bacterial SOS repair system, which is error-prone, and can induce the development of FQ resistance in *M. smegmatis* and *M. tuberculosis* with involvement of the SOS system.[383] Gatifloxacin (GTX) and moxifloxacin (MOX) appear to induce FQ resistant mutants at a lower rate than ciprofloxacin(CIP) and levofloxacin (LVF).[383] Furthermore GTX and MOX are more effective (lower MICs) than CIP and ofloxacin (OFX) and can be used to treat FQ low-level resistant isolates.[384] Von Groll *et al* has also observed that almost complete cross-resistance between OFX, MOX and GTX existed.[385]

The molecular basis of FQ resistance has been associated with *gyrA* and *gyrB* mutations in particular hotspots denominated Quinolone Resistance Determining Regions (QRDR).[386] GyrA QRDR is located between codons 74-95, whereas GyrB QRDR has been proposed to be comprehended between codons 500-540.[271, 387] Both QRDRs are located at the QBP according to the DNA gyrase crystal structure.[379, 387] The most common mutations associated with FQ resistance occur in *gyrA* in codons 94 and 90, in 46.2-71.9% and 4.0-38.5%, respectively.[219, 367, 380, 384, 385, 388-391] Mutations occurring in the *gyrB* gene have also been described, although with a lesser frequency and some with questionable role in FQ resistance.[367, 378, 392] Double mutations have also been described to act synergistically to confer FQ-resistance or, as associated with a reduced FQ susceptibility.[378, 390, 391]

Data from structural analysis show that *gyrA* and *gyrB* mutations can cause FQ resistance through the modification of the QBP geometry or lead to the modification of the DNA structure in the QBP when mutations in aminoacid residues that interact with the DNA occur.[379]

Regarding resistance levels, the wide range of MICs observed for a given mutation, different panels of tested FQs and, the diversity of methodologies employed in MIC determination hamper direct comparison.[384, 389-391] For instance, in a recent study, Cui *et al* have not found any statistical difference between MICs of strains with mutations on different *gyrA* codons or in *gyrB*.[389] Nevertheless several studies convey the notion that *gyrA* D94G/A/N mutations exhibit a higher resistance level than A90V and S91P, by this order.[384, 389-391] Differences in MIC-mutation correlations between these studies may also be related with the observation that FQ MIC can exhibit some degree of variability according with the strain's genetic background.[378]

Other mechanisms of FQ resistance might be involved as data from Singh *et al* showed that FQ efflux may play a role in FQ resistance.[393]

1.7 TB MOLECULAR EPIDEMIOLOGY

Molecular epidemiology can be defined as the study of the distribution and determinants of diseases that occur among human populations using molecular biology techniques.[394] The application of molecular methods in the genotyping of *M. tuberculosis* isolates has allowed obtaining more information on its propagation in miniepidemics or outbreaks.[394] Before molecular methods became available, strain discrimination relied on bacteriological data with reduced discriminatory power. The implementation of molecular data in the study of TB transmission, during the 80's, has allowed novel insights in TB epidemiology.[395]

The main principle behind TB molecular epidemiology relies in the assumption that high genotype variability exists among epidemiologically non-related *M. tuberculosis* isolates, while patients infected from a common source, or belonging to the same transmission chain, will have isolates with identical genotypes. Furthermore, isolates with identical or similar genotypes, *i.e.* clustered isolates, probably reflect cases that have arise due to recent transmission, whereas isolates with different genotypes reflect reactivation cases, of an infection acquired in the past, or even cases associated with migratory events.[59] Although the treatment of reactive cases or recent transmission cases is similar, recent transmission may be indicative of an outbreak or series of transmissions that require control measures in order to be contained.[395]

Genotypic data should, nevertheless, be interpreted at the light of the traditional epidemiologic information that is available and should consider the evaluation of a large percentage of the population's cases during a considerable time period.[59] This way, molecular epidemiology might provide additional data concerning TB transmission dynamics, distinguish exogenous reinfection from endogenous reactivation, determine the source of laboratory contaminations, identify risk factors for transmission in the community, and follow the geographical distribution and propagation of clones of public health concern.[59] Different reports have highlighted the importance of genotyping in improving traditional contact tracing in complex transmission networks where TB can be transmitted by brief contact and, in the identification of environments with increased risk for TB transmission.[396-399] Moreover,

molecular epidemiology is also invaluable in tracking the international dissemination and evolutionary changes in *M .tuberculosis* populations.[59, 394, 395]

Several genotyping methods using different genetic markers (*e.g.*, polymorphic GC-rich repetitive sequence, PGRS, or 16S and 23S rRNA-based methods) have been applied and tested for *M. tuberculosis* genotyping for the last three decades.[400] Nevertheless, only the three most widely used and consensual techniques employed in molecular epidemiology (Figure 1.12) are covered below. Still, when considering the choice of such method it is necessary to take in account that although a high discriminatory power is important to discriminate between epidemiologically non-related isolates, a discriminatory power too high can on the other end discriminate epidemiologically related isolates.[54, 395]

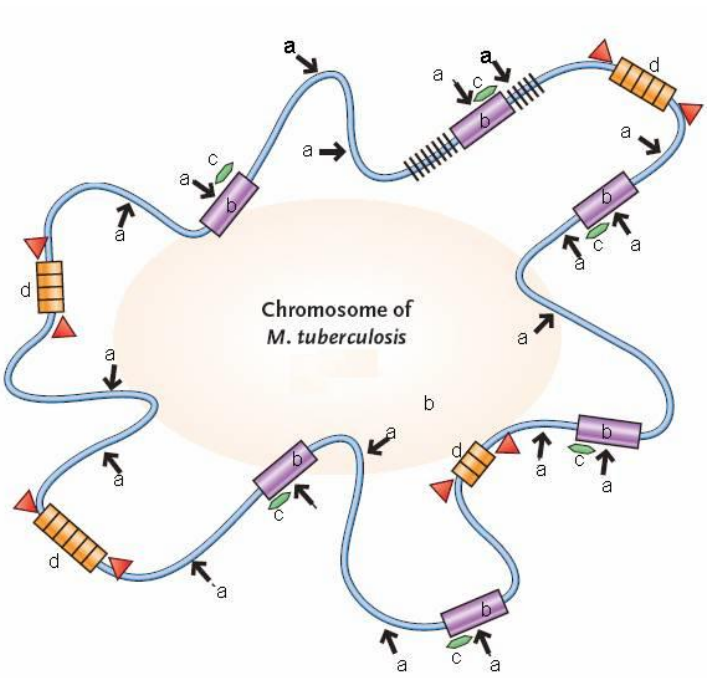


Figure 1.12 – *M. tuberculosis* chromosome representation and polymorphic *loci* used in molecular epidemiological methods. Legend: (a), *Pvu*II restriction sites; (b), IS*6110*; (c) IS*6110* probe; (d), MIRUs. Adapted from Barnes *et al*.[59]

One of the earliest typing techniques for TB was the Restriction Fragment Length Polymorphism (RFLP), with *Pvu*II, combined with a Southern hybridization using a partial sequence of the insertion sequence IS*6110* as a probe.[401] Insertion sequences are found dispersed throughout the bacterial chromosome in a variable number of copies and might be useful as probes to compare the number and size of the restriction fragment in which they can be found. IS*6110* has 1355 bp in length and is unique to the MTC.[402] Due to its polymorphic nature but relative stability in epidemiological related patients, the IS*6110* is also the genetic marker of choice for the RFLP-based typing method.[118, 401, 403, 404]

The RFLP-IS*6110* has a high discriminatory power but has, nevertheless, some drawbacks when compared with other available molecular typing techniques. RFLP-IS*6110* requires a considerably high amount of DNA (2 µg) that has to be extracted from cultured mycobacteria; for isolates with less than six IS*6110* copies, the discriminatory power becomes significantly

lower; it is of more difficult inter-laboratorial comparison; and, as it requires DNA extracted from culture it becomes a time-consuming technique.[59, 395]

RFLP-IS*6110* was employed in the first molecular epidemiologic studies carried out during the 90s in the Lisbon Health Region that have identified a MDR-TB outbreak occurring mainly among HIV seropositive patients.[26] Genotyping by RFLP-IS*6110* has also allowed the identification of a group of genetically close clusters that were first named as the A clusters and later, the Lisboa family.[26] Further cluster discrimination by *rpoB* mutations showed that independent transmission chains were linked with independent MDR-TB development followed by clonal expansion in the community.[405]

Another amply diffused typing method, because of its easy execution and because it is a PCR-based method, is Spoligotyping. This technique consists in the amplification of the spacer regions found within the DR (direct repeat) *locus*, wich is composed by 36 bp spaced tandem repeats. The methods consists in the amplification of these spacer regions and detection by reverse hybridization with specific probes for 43 spacer regions spotted onto a membrane followed by chemiluminescent detection.[406] This method will generate a hybridization profile for any clinical isolate. The spoligotyping profile can be expressed as a numerical profile (binary or octal code) which simplifies storing of typing data and greatly facilitates inter-laboratory comparison of typing data.[55, 60] Spoligotyping has nevertheless an inferior discriminatory power when compared with RFLP-IS*6110* which can lead to transmission overestimates becoming, however, useful in discriminating RFLP-IS*6110* clusters resulting from IS*6110* low copy-number strains.[407-409] Moreover, multiple and independent spacer and DR deletions have been reported to lead to convergent evolution.[410]

The Beijing family of strains was initially described among patients from People's Republic of China and Mongolia as a group of genetically close strains that shared 2/3 of the IS-containing *Pvu*II fragments but, having a identical spoligotype.[70] This highlights the usefulness of spoligotyping in tracking large and widespread genetic clades that exhibit some degree of divergence.

The most recent typing method is based on PCR amplification of different tandem repeat *loci* dispersed through the *M. tuberculosis* genome. These tandem repeat *loci* have a variable number of tandem repeats (VNTR) and are designated by Mycobacterial Interspersed Repetitive Unit (MIRU). As these *loci* are composed by small size tandem repeats (40-100 bp), the amplification and sizing of these *loci* allows the determination of the number of repeats present in each *locus* for each typed isolate, generating a numerical profile.[411, 412] The MIRU-

VNTR typing method started by including a set of 12 *loci*, which was subsequently increased up to 24 *loci* in order to achieve a higher discriminatory power. The discriminatory power of the 12-*loci* MIRU-VNTR has been reported as inferior to the one of RFLP-IS*6110* but, however, the 15 *loci* set has been proposed as equivalent to RFLP-IS*6110* and the 24-*loci* set as suitable for phylogenetic studies.[54, 56, 411] Additionally, MIRU-VNTR typing has the advantage of having a better discriminatory power when typing isolates with less than six IS*6110* copies, *e.g.*, *M. bovis* isolates. Other advantages of the MIRU-VNTR typing method consist in the rapid turn-around time and, on the fact that only about 30 ng of DNA is necessary for the complete 24-*loci* typing. Using fluorochrome-tagged primers it is possible to analyse multiplexed amplicons by capillary electrophoresis and automate the sizing and allele assigning processes.[57]

Nowadays, MIRU-VNTR typing has almost completely replaced RFLP-IS*6110* and it is considered the preferred typing method.

1.8 NEXT GENERATION SEQUENCING

The term "Next Generation Sequencing" (NGS) is commonly used to describe sequencing platforms based on more recent methodologies than the original sequencing based on Sanger chain termination chemistry or the chemical method from Maxam and Gilbert.[413, 414] While the original Sanger sequencing method required a specific primer to initiate reading through a template sequence, NGS strategy derives from the shotgun sequencing approach employed in the Human Genome Project.[414, 415] The new generation sequencing technologies are based on genomic fragmentation, adapter ligation and production of reads starting at random locations throughout the genome.[415] DNA templates are therefore sequenced in parallel in a stepwise iterative process in real-time in which each clonal template or single molecule is individually sequenced and analysed.[415] All NGS platforms end up by producing enormous amounts of data in a relatively short period and hence the reason that NGS is also associated with the term high-throughput sequencing.[416] Depending on the technology, the read length can vary as the first NGS technologies started by producing read lengths below the ones achieved by Sanger sequencing, although more recent technologies have surpassed the latter. Another important aspect to consider is coverage, *i.e.* the number of reads that overlap a given nucleotide position.

The first system to become available was from Roche, the GS-FLX 454 platform, and became available in 2004 using a combination of a sequencing-by-synthesis methodology known as

pyrosequencing and, single-molecule emulsion PCR.[417] Pyrosequencing is based on the detection of light from pyrophosphate groups released during dNTP incorporation.[415] The Illumina/Solexa Genome Analyzer was the second platform to become available and is presently the most widely used system.[415, 418] The Illumina platform uses a sequencing-by-synthesis methodology combining reversible chain termination and Bridge PCR for local cluster amplification.[415] Another widespread platform, the ABI SOLiD, uses a different approach for template sequencing by combining the emulsion PCR approach with a synthesis-by-ligation sequencing methodology using oligonucleotide probes.[415, 419]

These three platforms comprised the main second-generation sequencing technologies and are able to generate up to several billion bases per run.[416] Several updates have been made enabling higher read lengths.[420] Reads outputted by next generation sequencers can either be mapped to a reference genome, in a process known as reference assembly or, through finding overlapping regions be used without any reference genome in a *de novo* assembly process. Read length is an important factor as longer read lengths will in the future enable *de novo* assembly of more complex genomes.

A third generation of sequencing platforms is under development aimed at the development of sequencing methodology without using PCR for template amplification.[416] The third generation of sequencing technologies are based on a single-molecule sequencing system in which a read is derived from a single molecule and not from a clonal cluster.[416] Single-molecule systems have not only the advantage of minimizing errors that can be introduced by PCR but also to better reflect the relative abundance of template DNA. A shift in the detection method has also occurred among third-generation sequencers, from the more expensive optical imaging towards the detection by pH change in microstructures.[420] The Helicos HeliScope and the PacBio RS from Pacific Biosciences, are capable of single-molecule sequencing in real-time but still rely on fluorescence detection.[420-423] The Personal Genome Machine (PGM) sequencer from Ion Torrent/Life Technologies has successfully implemented single-molecule sequencing together with a non-optical detection system in a benchtop instrument.[416, 420] In the PGM, nucleotide incorporation is recorded by pH changes in a highly dense microwell array in which each well acts as a polymerization unit.[424] More recently, Ion Torrent/Life Technologies has released the Ion Proton Sequencer with a higher microwell density resulting in an increase in data output and allowing sequencing of a human genome on a chip.

The fourth generation of sequencers is expected to become available using nanopore technology, either biological or solid-state, that could generate long read lengths.[420] Solid-state

nanopores would be eventually more stable as it avoids the use of organic compounds.[420] Oxford Nanopore technologies are preparing to commercialize the first nanopore-based sequencer, the GridION system, that utilizes the heptameric protein α-hemolysin as bionanopores for strand sequencing.[425] This system will be based on threading single-stranded DNA through the nanopores as nucleotide specific voltage changes are registered in real-time.[420] Further information released by Oxford Nanopore technologies point towards read lengths in the order of tens of kb. Comparison between the above covered sequencing technologies is outlined in Table 1.7.

Table 1.7 – Comparison of the main NGS platforms representative of the different generations of sequencing technologies. Adapted from Niedringhaus *et al* [420] with updated information released by the manufacturers.

Generation	Company	Platform	Sequencing Method	Detection Method	Read Length (bp)
First	ABI/Life Technologies	3130xL – 3730xL	Capillary electrophoresis - Sanger	Fluorescence / Optical	600-1000
Second	Roche/454	GS FLX Tiatanium XL+ System	Pyrosequencing, sequencing by synthesis	Optical	Up to 1000
Second	Illumina	HiSeq2500	Reversible terminator, sequencing by synthesis	Fluorescence / Optical	2 x 100 (2 x 150 in Rapid Run Mode)
Second	ABI/SOLiD	5500xl W SOLiD System	Sequencing by ligation	Fluorescence / Optical	50 - 75
Second	Helicos	HeliScope	Single-molecule sequencing by synthesis	Fluorescence / Optical	25-35
Third	Pacific Biosciences	PacBio RS	Real-time, single-molecule DNA sequencing	Fluorescence / Optical	Up to 20 Kb (average 3-5 Kb)
Third	Ion Torrent / Life Technologies	Personal Genome Machine (PGM) sequencer	Sequencing by synthesis	pH change	Up to 400
Third	Ion Torrent / Life Technologies	Ion Proton	Sequencing by synthesis	pH change	Up to 200
Fourth	Oxford Nanopore	gridION	Nanopore strand sequencing	Electric Current	unknown

The use of NGS technologies to study the genomic diversity of *M. tuberculosis* is producing interesting data on the diversity, mode of evolution and host adaptation. Some studies have

employed the capability of next generation sequencing in studying the microevolution of *M. tuberculosis* in specific settings.[426, 427] Schurch *et al* have conducted a genome-wide analysis of two isolates, representing the first and the end isolate from a well studied transmission chain, and have found six different polymorphisms and an IS*6110* transposition event. Further isolate characterization has suggested that the molecular evolution of *M. tuberculosis* in vivo is characterized by periods of relative genomic stability, followed by bursts of mutations[428]. Ford *et al* using the cynomolgous macaque model of infection studied the mutation rate of *M. tuberculosis* during latency by NGS found that *M. tuberculosis* acquires a similar number of chromosomal mutations during latency as occurs during active disease or in a logarithmically growing culture over the same period of time. [429]

Additionally, the ability to compare drug-resistant and susceptible strains with the same genotype identifies not only resistance-associated mutations but might also highlight compensatory mutations or others important in the process of drug resistance acquisition and physiological adaptation. [430]

1.9 OBJECTIVES

In Portugal, as already stated, TB constitutes a major public health concern, more specifically MDR-TB. It is in the two largest urban centers (Lisbon and Oporto) that this infectious disease associated with drug resistance phenotypes exerts its highest impact nationwide. In Lisbon, early studies carried out during the 1990s have shown that this situation is mostly due to the circulation of genetically close strains: the Lisboa family. An increasing knowledge of the molecular aspects of TB epidemiology can contribute to a better understanding of the dynamics of TB transmission and drug resistance development. This work is therefore based and covers the following:

- Genotypic analysis and molecular basis of resistance in MDR-TB isolates

In 2006, the designation of XDR-TB came into place to name what is now known to be an old problem associated with TB resistance to second-line drugs but, simultaneously imposed another question regarding TB control in Portugal: what is the prevalence of XDR-TB in the country and in the region?[431, 432] Moreover, as MIRU-VNTR genotyping method became available it also became important an updated genotypic analysis of MDR-TB clinical isolates and drug-resistance associated mutations.[57, 412, 433] The prevalence of specific drug resistance

mutations is of special importance for the implementation or development of rapid molecular tests. These research questions are addressed in Chapter 2 through a MIRU-VNTR genotypic analysis of MDR-TB strains recovered from Lisbon Health Region within a single year (2003) with the purpose of assessing the prevalence and clonality of circulating MDR-TB strains.[24]

Another important aspect that is focused throughout this work concerns the molecular basis of resistance. Since the prevalence of drug-resistance conferring mutations appear to vary geographically, it is highly important to know which mutations are responsible for drug resistance so that molecular resistance testing may be implemented in a setting-adapted manner. In Chapter 2 the main genes associated with resistance to first-line drugs were screened for mutations in order to establish the most prevalent mutations associated with drug resistance.

- Molecular determinants of resistance to second-line drugs and diversity of XDR-TB isolates

Following the previous line of research and given the high XDR-TB rate found previously, the analysis of genes associated with resistance to second-line drugs was also found to be important. The screening of *gyrA*, *rrs* and *tlyA* genes from XDR-TB isolates allied with MIRU-VNTR genotyping is described in Chapter 3.[25] Using this approach we intend to identify the major XDR-TB genetic clusters in Lisbon Health region as well as finding drug-resistant associated mutations that can be used as molecular markers for specific clusters.

- Laboratory perspective of the M/XDR-TB situation in Lisbon

Another objective, here delineated in the sequence of the previous studies, is to provide further evidence for a continued endemic M/XDR-TB situation in Lisbon, from laboratory data. In Chapter 4 the results from the compiled the records of 3 025 *M. tuberculosis* clinical isolates recovered between 2001-2006 are analysed in combination with MIRU-VNTR genotyping data from 100 isolates resistant to at least one first-line drug.

- Cross-resistance between second-line aminoglycosides

The lack of an apparent resistance determinant to KAN/AMK in Chapter 3 and a more recent report by Zaunbrecher *et al*[372] that associated *eis* gene overexpression with KAN resistance in *M. tuberculosis* reopened the question of the molecular basis of KAN resistance that we were unable to answer previously.[434] With the objective of gaining insight on the KAN/AMK cross-resistance mechanism we sought to characterize the promoter region of the *eis* gene in order

to find mutations that may exert a putative upregulatory effect in the strains previously tested and, strains isolated during the 90s outbreak known to belong to the Lisboa3 cluster (Chapter 5). Another aim of this study is to correlate hypothetical mutations to be found in the promoter region of the *eis* gene with the AMK/KAN resistance level.

- Contribution of *gidB* mutations to STP resistance and phylogeny

Given previously associations between *gidB* mutations and STP-resistance, it would be of interest to screen this gene for mutations that could correlate with STP-resistant isolates with no *rrs* or *rpsL* mutations and investigate the resistance levels.[325, 326, 435] The role of *gidB* gene and its contribution for STP resistance level by screening of STP-resistant and –susceptible isolates is addressed in Chapter 6. As *gidB* gene mutations have been proposed to be associated with specific phylogenetic lineages, any possible mutation-cluster association will also be investigated.

- Genomic distinctiveness and diversity of M. tuberculosis isolates in Lisbon, Portugal

The two major clusters associated with MDR-TB in Lisbon Health Region appear to be highly endemic to Portugal and, unique from a genotyping standpoint. The Lisboa family and the Q1 strains appear to show a high degree of success for transmission in the community, resulting in a high local prevalence (Chapters 2-4). Several questions emerge regarding these strains:

1. What is the phylogenetic positioning and/or phylogeographic origin of both strain types?
2. Do these strains possess any virulence factors already known or yet to discover?
3. What is the nature of its association with drug resistance and what are the microevolutionary events accompanying drug-resistance acquisition?

In an attempt to answer these questions and investigate the genomic diversity present in the MDR-TB strains circulating in the region, we sought to perform a comparative genomic analysis encompassing these strains and others circulating in Lisbon Health Region. Through a collaboration with the London School of Hygiene and Tropical Medicine and the King Abdullah University of Science and Technology, the genomes of 56 clinical isolates were sequenced by NGS technology, the great majority of which recovered in Lisbon Health Region. The first results from this comparative genomic analysis are described and discussed in Chapter 7.

Globally, all the data coming from these more concrete objectives and studies will contribute to a main purpose of gaining a better picture of MDR-TB epidemiology in Lisbon and towards

M. tuberculosis biology and adaptation in this host population. This approach can ultimately contribute to the deployment of effective measures to control MDR-TB.

CHAPTER 2

Multidrug-Resistant Tuberculosis in Lisbon, Portugal: A Molecular Epidemiological Perspective

João Perdigão[1], Rita Macedo[2], Inês João[2], Elisabete Fernandes[2], Laura Brum[2] and Isabel Portugal[1,2]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Unidade de Micobactérias, Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal

Microbial Drug Resistance 14(2) 133-43

2.1 ABSTRACT

Portugal has the fourth highest TB incidence rate of the European Union. Thirty-nine percent of all cases originate in Lisbon Health Region. Portugal also presents high levels of MDR-TB (1.1%, primary rate and 6.7%, in retreatment cases). In the present study we have characterized 58 MDR-TB clinical isolates by: i) determining the resistance profile to first and second-line drugs used in the treatment of tuberculosis; ii) genotyping all isolates by MIRU-VNTR; iii) analyzing mutations conferring resistance to INH, RIF, STP and PZA, in *katG*, *mabA-inhA*, *rpoB*, *rpsL*, *rrs* and *pncA* genes. We have therefore established the prevalence of the most common mutations associated with drug resistance in the Lisbon Health Region; C-15T in *mabA-inhA* for INH; S531L in *rpoB* for RIF; K43R in *rpsL* for STP; and V125G in *pncA* for PZA. By genotyping all isolates and combining with the mutational results we were able to assess the isolates genetic relatedness and determine possible transmission events. Strains belonging to family Lisboa, characterized several years ago, are still responsible for the majority of the MDR-TB. Even more alarming is the high prevalence of extensive drug resistant tuberculosis (XDR-TB) among the MDR-TB isolates, which was found to be 56%. The TB status in Portugal therefore requires urgent attention in order to contain the strains continuously responsible for MDR-TB and now, XDR-TB.

2.2 INTRODUCTION

Portugal has the fourth highest tuberculosis (TB) incidence rate in the European Union (EU), having registered in 2004, 33.7 cases per 100 000 habitants nationwide. Although this rate has diminished in the last 20 years to less than a half, the decline rate has been very slow, having decreased only 2.7 cases per 100 000 people from 2003 to 2004.[436]

In the year of 2004, 39% of all new cases reported nationwide originated from Lisbon Health Region. The majority of these cases (72%) were concentrated in Lisbon district, of which 74% were associated with three high-risk groups for TB, namely, human immunodeficiency virus (HIV) infected, drug users and immigrants.[436]

The emergence of MDR-TB, that is, resistance to at least INH and RIF, poses a serious threat, not only to national tuberculosis control plans, but also to the implementation of an effective directly observed short-course treatment (DOTS) strategy. Portugal also reports high rates of MDR-TB. In 2004, Portuguese Health authorities have reported a rate of primary MDR-TB of 1.1%, and 6.7% among retreatment cases.[436]

Drug resistance generally occurs as a result of a regimen with the wrong combination of drugs or, due to the patient lack of adherence to the regimen. The only known mechanism for resistance development by *M. tuberculosis*, is through mutation acquisition in chromosomal genes associated with drug resistance. Several genes have been associated with drug resistance, namely *katG*, *inhA*, *kasA* and *ahpC*, with INH;[207, 437, 438] *rpoB* with RIF resistance;[250, 438, 439] *rpsL* and *rrs* with STP resistance;[311, 313, 440] *embCAB* with EMB resistance;[278, 284, 289] and *pncA* with PZA resistance.[332, 336, 340, 347, 441] The emergence of MDR-TB strains is due to sequential acquisition of mutations in some of the genes above.

Epidemiological surveys concerning MDR-TB strains are necessary to gain insight into the transmission and resistance development mode of TB, and ultimately to manage properly the local TB situation. Strain typing is an invaluable process in epidemiology and, the last years epidemiological surveys have benefited from the development of additional molecular epidemiological methods, alternative to the traditional Restriction Fragment Length Polymorphism (RFLP) followed by hybridization with Insertion Sequence (IS) *6110*.[59, 403] Such an example is the use of MIRU-VNTR for strain typing. MIRU-VNTR typing is based on the determination of the number of repeating units on several chromosomally dispersed *loci*.[57, 59, 394] Although the discriminative power reported by some studies is lower than that of RFLP-

IS*6110*, it is not as time-consuming as the former, facilitates inter-laboratorial data exchange, and is less likely to underestimate the recentness of transmission, as reported for RFLP-IS*6110*.[54, 442]

In the present study, we have characterized several resistance-associated genes from 58 MDR-TB isolates recovered from Lisbon Health Region and, assessed the genetic relatedness of these isolates by MIRU-VNTR. As the emergence of more deadly strains of *M. tuberculosis* occur throughout the world we have assessed the predominance of strains with extensive drug resistance (XDR-TB, *i.e*., MDR-TB plus resistance to at least one injectable second-line drug and a FQ) among the studied isolates.[432, 443, 444] Since mutations conferring resistance varies geographically, the characterization of the most prevalent mutations in a given community allows the development of locally adapted, resistance molecular screening tests. Allying the determination of resistance-associated mutations with strain typing by MIRU-VNTR not only allows us to gain knowledge concerning the mode of resistance acquisition, but also permit a more accurate characterization of TB transmission mode in the community, and the identification of possible MDR-TB outbreaks.

2.3 MATERIALS AND METHODS

Clinical Isolates and DNA

Fifty-eight MDR-TB isolates were recovered from patients whose clinical samples were sent for testing at the mycobacteriology unit at the National Institute of Health Dr. Ricardo Jorge (INSA-RJ), Lisbon. The isolates were collected during the year of 2003 and are a subset of 116 MDR-TB strains isolated in the same year. Isolates were cultured on LJ slants or, in the BACTEC 960 (Becton Dickinson™) system and then passed to LJ slants. Identification of *M. tuberculosis* complex was accomplished using the Accuprobe (Gen-Probe™) method.

DNA of the above isolates was extracted as previously described by van Soolingen *et al*.[445]

Isolates were tested for first-line drug susceptibility by the BACTEC 960™ MGIT™ methodology for INH (0.1 mg/L), RIF (1.0 mg/L), STP (1.0 mg/L), EMB (5.0 mg/L) and PZA (100 mg/L), according to the manufacturer‘s instructions. Isolates resistant to at least INH and RIF were considered MDR-TB and, tested for second-line drug susceptibility by the radiometric BACTEC

460™ system to CAP (1.25 mg/L), AMK (1.0 mg/L), KAN (5.0 mg/L), CIP (1.0 mg/L), OFX (2.0 mg/L), ETH (5.0 mg/L) and PAS (4.0 mg/L), according to Pfyffer *et al*.[446]

MICs

PZA minimum inhibitory concentration was determined using concentrations of 100, 300 and 900 mg/L by the BACTEC 960 MGIT system. Minimum inhibitory concentrations were defined as, the minimum tested drug concentration that rendered a susceptible result.

Gene amplification and sequencing

Several resistance associated genes from MDR-TB clinical isolates have been amplified and characterized according to the respective resistance pattern. We have characterized the *katG* gene and the *mabA-inhA* operon regulatory region, in clinical isolates resistant to INH; *rpoB* resistance determinant region (RRDR) in RIF resistant isolates; *rpsL* and *rrs* for STP resistant isolates; and *pncA* for PZA resistant isolates. All amplification reactions were carried out using PuReTaq™ Ready-To-Go™ PCR beads (GE Healthcare™), with 10 pmol of each oligonucleotide primer used and approximately 50 ng of template DNA, accordingly to the manufacturer's instructions.

A 620 bp internal fragment (904-1523) of *katG* gene was amplified by polymerase chain reaction (PCR) with forward oligonucleonide primer katG904 and reverse oligonucleotide primer katG1523 (Table 2.1). The cycling conditions consisted in an initial denaturing step of two minutes at 94ºC, followed by 50 cycles of amplification (denaturation at 94ºC for 1min; annealing at 60ºC for 1 min; and extension at 72ºC for 1 min).

The *mabA-inhA* regulatory region was amplified in a 248 bp fragment (positions -168 to 80, relative to *mabA-inhA* initiation codon) using forward oligonucleotide primer mabA-inhA-F and reverse oligonucleotide primer mabA-inhA-R (Table 2.1). Cycling conditions for this amplification consisted in an initial 5 min denaturation step; followed by 30 cycles of denaturation at 94ºC for 5 min, annealing at 65ºC for 1 min, and an extension step at 72ºC for 1,5 min; a final extension step at 72ºC for 10 min was performed.

A 409 bp fragment of *rpoB* (GenBank accession no. L27989) encompassing the RIF resistance-determinating region was amplified by PCR with forward oligonucleotide primer TR1 and, reverse oligonucleotide primer RPOB-2 (Table 2.1). The cycling conditions consisted in a denaturation step at 94ºC for 4 min, followed by 30 cycles of amplification (denaturation at

94ºC for 1 min, primer annealing at 60ºC for 1 min, and extension at 72ºC for 1 min), and a last final extension step at 72ºC for 15 min.

Table 2.1– Oligonucleotide primers used to amplify and sequence the studied genes.

	Sequence	Tm (ºC)	Reference
katG904	5'-AGCTCGTATGGCACCGGAAC-3'	68,4	Uhl *et al* [447]
katG1523	5'-TTGACCTCCCACCCGACTTG-3'	69,5	Uhl *et al* [447]
mabA-inhA-F	5'-CCTCGCTGCCCAGAAAGGGA-3'	66,6	Coll *et al* [448]
mabA-inhA-R	5'-ATCCCCCGGTTTCCTCCGGT-3'	66,6	Coll *et al* [448]
PNCA-2R	5'-GAACACCGCCTCGATTGCCG-3'	74.0	Hannan *et al* [449]
PNCA-8	5'-GGTTGGGTGGCCGCCGGTCAG-3'	80.2	Hannan *et al* [449]
PNCA-6	5'-CCTCGTCGTGGCCACCGC-3'	75.0	Hannan *et al* [449]
PNCA-10	5'-GCTGGTCATGTTCGCGATCG-3'	71,3	Hannan *et al* [449]
PNCA-11	5'-GCTTTGCGGCGAGCGCTCCA-3'	78.6	Hannan *et al* [449]
RPOB-1	5'-GGGAGCGGATACCACCC-3'	73.0	Portugal *et al* [405]
RPOB-2	5'-GCGGTACGGCGTTTCGAT-3'	73.1	Portugal *et al* [405]
RPSL-1	5'-GGCCGACAAACAGAACGT-3'	63.8	Kirschner *et al* [450]
RPSL-2	5'-GTTCACCAACTGGGTGAC-3'	58.8	Kirschner *et al* [450]
RRS-1	5'-GAGAGTTTGATCCTGGCT-3'	62.9	Kirschner *et al* [450]
RRS-2	5'-TGCACACAGGCCACAAGG-3'	72.5	Kirschner *et al* [450]
TR1	5'-TACGGTCGGCGAGCTGAT-3'	71.7	Portugal *et al* [405]

The entire open reading frame (ORF) of *rpsL* (GenBank accession no. L08011.1), plus 46 nucleotides upstream and 86 nucleotides downstream, were amplified by PCR in a 504 bp fragment using forward oligonucleotide primer RPSL-1, and reverse oligonucleotide primer RPSL-2 (Table 2.1). Cycling conditions for this amplification consisted in 40 cycles of a denaturation step at 94ºC for 1 min, a primer annealing step at 57ºC for 2 min and, an extension step at 72ºC for 2 min.

The region comprehended by nucleotides 10 to 1037 in *rrs* was amplified by PCR using forward oligonucleotide primer RRS-1 and reverse oligonucleotide primer RRS-2 (Table 2.1). The cycling conditions consisted in 40 cycles of a denaturation step at 94ºC for 1 min, a primer annealing step at 60ºC for 2 min and, an extension step at 72ºC for 2 min.

The entire *pncA* ORF, as well as 124 bp upstream and 59 bp downstream, was amplified in a 744 bp fragment by PCR, using forward oligonucleotide primer PNCA-11 and, reverse oligonucleotide primer PNCA-8 (Table 2.1). Cycling conditions consisted in: a denaturation step at 94ºC for 5 min; 20 cycles of denaturation at 94ºC for 30 sec, primer annealing at 68ºC for 30 sec, with a decrease of 0.5ºC per cycle and, extension at 72ºC for 30 sec; 15 cycles of

denaturation at 94ºC for 30 sec, primer annealing at 58ºC for 30 sec and, extension at 72ºC for 30 sec; finally an extension step at 72ºC for 8 min.

Amplicons were purified with MicroSpin™ S-300 HR Columns (GE Healthcare™) prior to sequencing. Sequencing reaction was performed with BigDye™ Terminator Cycle Sequencing Kit with AmpliTaq™ DNA polymerase (Applied Biosystems™) using oligonucleotide primers katG904 for katG amplicons; mabA-inhA-F for *mabA-inhA* regulatory region amplicons; RPOB-1 for *rpoB* amplicons; RPSL-1 for *rpsL* amplicons; RRS-1 and RRS-2 for *rrs* amplicons; and, PNCA-6, PNCA-10, PNCA-11 and PNCA-2R for *pncA* amplicons (Table 2.1). Obtained sequences, were aligned with the respective wild-type alleles, from *M. tuberculosis* H37Rv, using CLC Free Workbench (CLC Bio™, Aarhus C, Denmark) and analyzed using BioEdit (v.7.0.5.2, T.A. Hall) software.

Endonuclease digestions

To rapidly detect some resistance-associated mutations, RFLP analysis was used. Mutations S315T and R463L, in *katG*, were detected by *Msp*I (New England Biolabs™) digestion of the *katG* amplicon. Two micrograms of PCR product were digested with 10U of *Msp*I in accordance with the manufacturer's instructions for 1 h at 37ºC. The analysis of the restriction profile allows detection of mutations G944C (S315T) and G1388T (R463L) as described by Uhl *et al*.[447]

Mutation A130G in *rpsL* was detected by *Mbo*II (New England Biolabs™) digesting of the *rpsL* amplicon. Two micrograms of PCR product were digested with 5U of *Mbo*II accordingly to the manufacturer's instructions for 1 h at 37ºC. Absence of restriction indicates the presence of mutation A130G.

Mutations T359C and T374G in *pncA* were detected by *Sac*II and *Btg*ZI (New England Biolabs™) digestion, respectively. Digestions were done independently and sequentially using approximately 1 µg of PCR product, with 10 U of *Sac*II or with 1 U of *Btg*ZI, for 1 h at 37ºC for *Sac*II, or 60ºC for *Btg*ZI as recommended by the manufacturer. Occurrence of restriction indicates the presence of the tested mutation.

MIRU-VNTR genotyping

MIRU-VNTR genotyping was performed for each isolate by multiplex PCR amplification of 12 MIRU-VNTR *loci* using HotStarTaq™ DNA polymerase (Qiagen™), as previously described by Supply *et al*.[57]

Clustering analysis

Isolates were clustered accordingly to their MIRU-VNTR profile using BioNumerics software (Applied Maths™, Sint-Martens-Latem, Belgium). Pearson coefficient was used to calculate the similarity matrix and clustering was performed by the unweighted pair group method with arithmetic mean (UPGMA).

2.4 RESULTS

Multidrug resistance. In 2003, 116 MDR-TB isolates corresponding to the same number of patients, from 22 different hospital units across Lisbon's Heath Region, were isolated or received for testing at the Portuguese National Institute of Health – Dr. Ricardo Jorge's Unit of Mycobacteriology. Of these 116 isolates, 58 have been fully characterized since the remaining isolates had lost viability. These 58 isolates originated from 19 different heath institutions units across Lisbon's Health Region (Figure 2.1), herein designated as health institutions A to S. Among these isolates, six different resistance patterns to first-line antibacillary drugs were observed (Table 2.2); resistance to all first-line antibacillary drugs was the most common pattern.

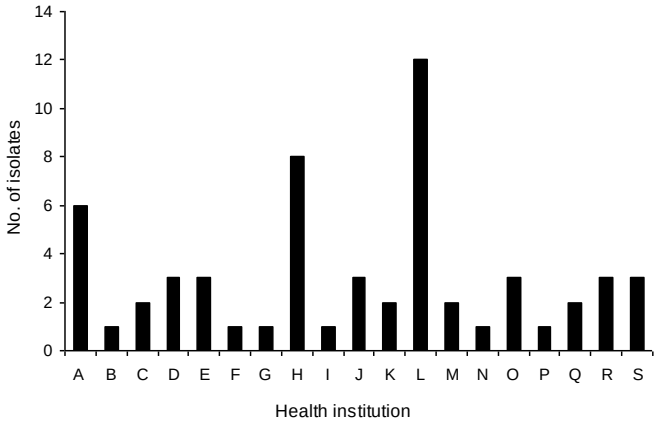


Figure 2.1 - Distribution of the studied isolates through the original health institutions here designated as A to S.

Table 2.2 – Resistance patterns to first-line drugs observed in the studied isolates, I, INH; R, rifampicin; S, streptomycin; E, ethambutol; P, pyrazinamide.

Resistance pattern	No. of isolates (%)
IR	2 (3.4)
IRP	5 (8.6)
IRS	5 (8.6)
IRSE	3 (5.2)
IRSP	16 (27.6)
IRSEP	27 (46.6)
Total	**58**

Resistance Associated Mutations. We started the characterization of the 58 MDR-TB isolates by determining the molecular basis of resistance, accordingly to the resistance pattern exhibited by each. INH-resistant isolates were screened for mutations for mutations at the *katG* gene and *mabA-inhA* regulatory region; RIF resistance-determining region (RRDR) of *rpoB* gene was characterized by sequencing in RIF-resistant isolates; *rpsL* gene and STP resistance-determining region in *rrs* gene were characterized by PCR-RFLP and sequencing in STP-resistant isolates; finally, the *pncA* gene and its putative promoter have been characterized by PCR-RFLP and sequencing in PZA-resistant isolates.

In the characterization of resistance to INH we started by characterizing an internal region of *katG* gene, associated with resistance in previous studies.[437, 448] We found that seven isolates displayed a serine-to-threonine substitution at codon 315, while the remaining isolates did not exhibit any mutation at the referred region (Table 2.3). All 58 isolates were also screened for mutations in the *mabA-inhA* regulatory region; 53 (91.4%) isolates had a cytosine-to-thymine substitution at position -15, and the remaining five did not exhibit any mutation in the same region.

Mutations conferring resistance to RIF were found in 57 (98.3%) out of 58 isolates resistant to RIF. We have found only 2 different mutations in the RRDR (Table 2.3), all of which previously associated with RIF resistance. Fifty-three isolates had a serine-to-leucine substitution at codon 531 and 4 isolates had an aspartic acid-to-valine substitution at codon 516 of RpoB protein.

The screening for mutations conferring resistance to STP in *rpsL* and *rrs* among the 51 STP-resistant isolates, resulted in the identification of only one mutation – a lysine-to-arginine change at codon 43 of the RpsL protein, detected in 34 STP-resistant isolates (Table 2.3). The remaining 17 (33,3%) isolates didn't have any mutation at the *rpsL* gene or at the *rrs* STP resistance-determining region.

Twelve different mutations in the *pncA* gene were found among the 48 PZA-resistant isolates, the most common being a thymine-to-cytosine and a thymine-to-guanine substitution at nucleotides 359 and 374, respectively (Table 2.3). Three new mutations were identified in the course of this study, a thymine insertion at nucleotide position 282; a 9 bp deletion at nucleotide position 381; and a cytosine-guanine insertion at nucleotide position 439. All the PZA-resistant isolates had a mutation in the *pncA* gene. Mutations 282InsT and 381Del9bp were associated with medium- (>300 mg/L) and high- (>900 mg/L) level resistance, respectively.

Table 2.3 – Mutations found in resistance-associated genes of the 58 studied clinical isolates.

Gene	Mutation (nt)	Mutation (aa)	No. of Isolates (%)
katG			
	None	None	51(87.9)
	G944C	S315T	7(12.1)
		Total	**58**
inhA			
	None	None	5(8.6)
	C-15T		53(91.4)
		Total	**58**
rpoB			
	None	None	1(1.7)
	C2431T	S531L	53(91.4)
	A2386T	D516V	4(6.9)
		Total	**58**
rpsL			
	None	None	17(33.3)
	A130G	K43R	34(66.7)
		Total	**51**
pncA			
	T2C	M1T	7(14.6)
	A226C	T76P	1(2.1)
	250InsC	Frameshift	3(6.3)
	282InsT	Frameshift	1(2.1)
	C285A	Stop	1(2.1)
	T359C	L120P	11(22.9)
	T374G	V125G	19(39.6)
	381Del9bp	Frameshift	1(2.1)
	392InsG	Frameshift	1(2.1)
	T398C	I133T	1(2.1)
	T416G	V139G	1(2.1)
	439InsCG	Frameshift	1(2.1)
		Total	**48**

MIRU-VNTR Clusters. All 58 isolates were genotyped by MIRU-VNTR and grouped in clusters, that is, isolates with 100% similarity index, accordingly to the similarity of their MIRU-VNTR profile. Forty-three (74.1%) isolates were grouped in nine clusters, which suggests a recent high-transmission rate (Figure 2.2). Families were designated and named accordingly to previous Lisbon Health Region's RFLP-IS*6110* typing results.[26]

The most predominant family is family Lisboa, initially identified by Portugal *et al*[26] as Clusters A, containing 32 (55.2%) isolates and 5 of the 9 clusters identified in this study.

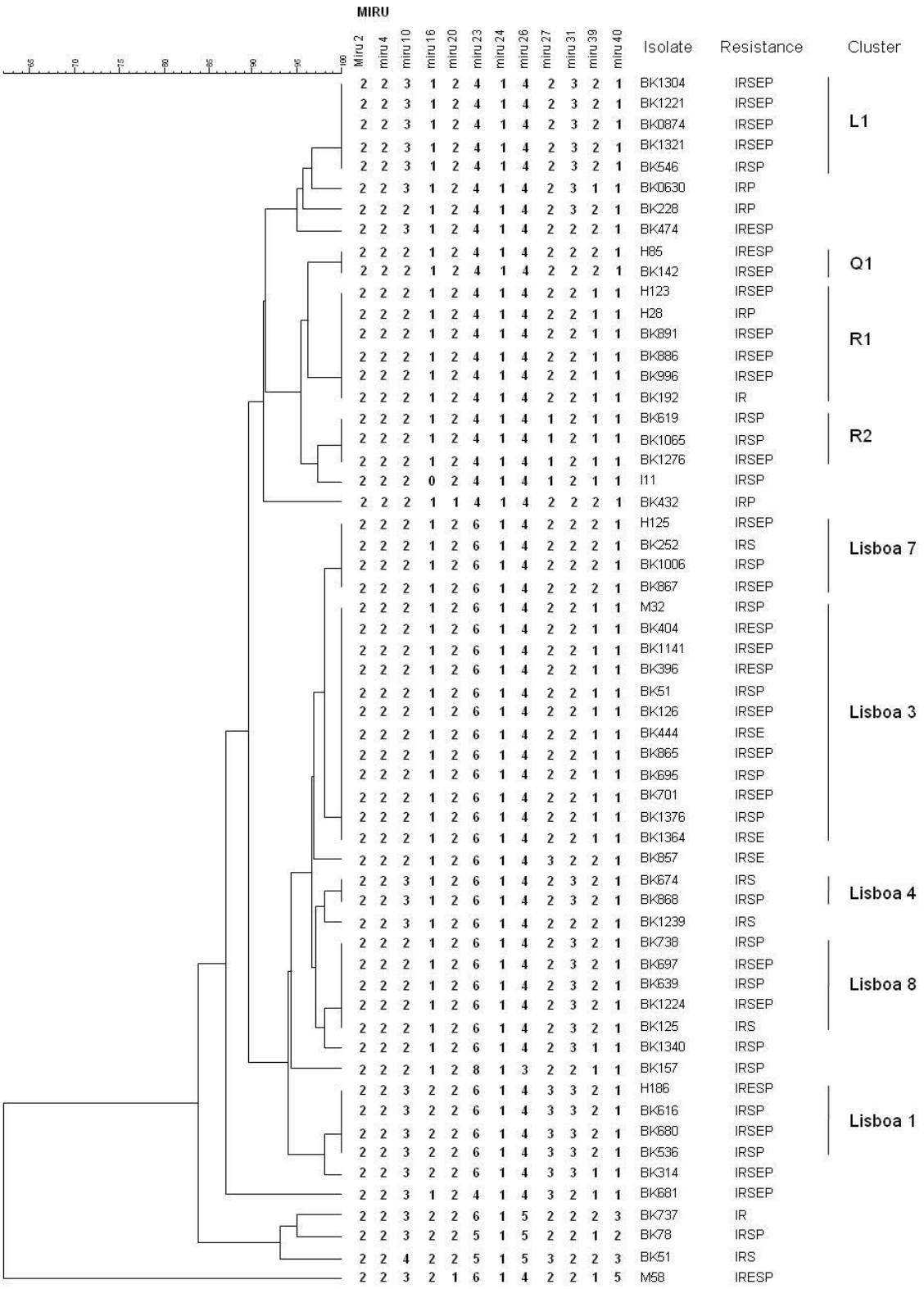


Figure 2.2 – MIRU-VNTR dendrogram of the isolates characterized in the present study. Resistance profile: I, isoniazid; R, rifampicin; S, streptomycin; E, ethambutol; P, pyrazinamide**.**

One of the clusters of the Lisboa family, named Lisboa3, has also been identified in previous studies by some of us as the most predominant cluster, which allowed the detection of at least two independent outbreaks.[405] In this study, Lisboa3 remains the most predominant cluster of isolates: 12 (20.7%) isolates originating from 10 different hospital units belong to this cluster.

Cluster's mutational analysis. Combining a cluster analysis with a mutational analysis of a given cluster allows the discrimination of that cluster, permitting the differentiation of independent and more recent transmission events.[405, 451] Results are presented in Table 2.4.

In cluster R2 one of the isolates differs from the remaining isolate by having additional resistance to EMB; such, points to a common origin but a new strain has already arose in this cluster. Cluster R1 although constituted by 6 clinical isolates has 5 different strains. Isolates H123 and BK891 probably represent the same strain since they share the same resistance and mutational profile. Cluster Q1 is constituted by only 2 isolates that probably represent the same strain. Cluster L1 has one different isolate that acquired resistance to EMB. Cluster Lisboa7 isolates correspond to different strains since there isn't a common resistance and mutational profile. Cluster Lisboa3 is composed by 12 clinical isolates from 10 different hospital units. Several strains are represented in this cluster; isolates BK404, BK1141, BK126 and BK1376 share the same mutational profile, isolate BK1376 differs from the remaining by lacking resistance to EMB. The remaining Lisboa clusters present similar situations as the Lisboa7 cluster, in which each isolate represents a different strain since there is no common mutational and resistance profile among isolates.

Nosocomial transmission does not seem to be significant, as only two identical isolates in one hospital and another two in another hospital were detected.

Analysing the mutations present in a given cluster, also allows an insight on the mode of resistance acquisition in the given family or cluster. We have found that mutation K43R in RpsL protein is present in all isolates of Lisboa family. Analysing Lisboa3 cluster we can observe that, probably, resistance to STP was the first to be acquired, since mutation K43R is common to all isolates, followed by acquisition of resistance to RIF, then INH and last to PZA or EMB.

Table 2.4 – Clustered isolates mutational and origin characterization.

Cluster	Isolate	Origin	Resistance	InhA	KatG	RpoB	RpsL	PncA
R2	BK619	L	IRSP	C-15T	none	S531L	none	T374G
R2	BK1065	E	IRSP	C-15T	none	S531L	none	T374G
R2	BK1276	E	IRSEP	C-15T	none	S531L	none	T374G
R1	H123	H	IRSEP	C-15T	none	S531L	none	T374G
R1	H228	H	IRP	C-15T	none	S531L	-	T374G
R1	BK891	E	IRSEP	C-15T	none	S531L	none	T374G
R1	BK886	R	IRSEP	C-15T	none	none	none	T374G
R1	BK996	D	IRSEP	C-15T	none	S531L	K43R	T359C
R1	BK192	J	IR	none	S315T	S531L	-	-
Q1	H85	H	IRSEP	C-15T	none	S531L	none	T374G
Q1	BK142	H	IRSEP	C-15T	none	S531L	none	T374G
L1	BK1304	A	IRSEP	C-15T	none	S531L	none	T374G
L1	BK1221	A	IRSEP	C-15T	none	S531L	none	T374G
L1	BK874	J	IRSEP	C-15T	none	S531L	none	T374G
L1	BK1321	R	IRSEP	C-15T	none	S531L	none	T374G
L1	BK546	R	IRSP	C-15T	none	S531L	none	T374G
Lisboa7	H125	H	IRSEP	C-15T	none	S531L	K43R	439InsCG
Lisboa7	BK252	Q	IRS	C-15T	none	S531L	K43R	-
Lisboa7	BK1006	S	IRSP	C-15T	none	S531L	K43R	T359C
Lisboa7	BK867	L	IRSEP	C-15T	none	S531L	K43R	T359C
Lisboa3	M32	M	IRSP	C-15T	S315T	D516V	K43R	250InsC
Lisboa3	BK404	H	IRSEP	C-15T	none	S531L	K43R	T2C
Lisboa3	BK1141	A	IRSEP	C-15T	none	S531L	K43R	T2C
Lisboa3	BK396	L	IRSEP	None	S315T	S531L	K43R	T416G
Lisboa3	BK51	P	IRSP	C-15T	none	S531L	K43R	T359C
Lisboa3	BK126	K	IRSEP	C-15T	none	S531L	K43R	T2C
Lisboa3	BK444	F	IRSE	C-15T	none	S531L	K43R	-
Lisboa3	BK865	L	IRSEP	C-15T	S315T	S531L	K43R	T359C
Lisboa3	BK695	B	IRSP	C-15T	none	S531L	K43R	282InsT
Lisboa3	BK701	L	IRSEP	C-15T	none	S531L	K43R	T359C
Lisboa3	BK1376	O	IRSP	C-15T	none	S531L	K43R	T2C
Lisboa3	BK1364	N	IRSE	C-15T	none	S531L	K43R	-
Lisboa8	BK738	L	IRSP	C-15T	none	S531L	K43R	T2C
Lisboa8	BK697	S	IRSEP	C-15T	none	S531L	K43R	T359C
Lisboa8	BK639	S	IRSP	C-15T	none	S531L	K43R	T374G
Lisboa8	BK1224	D	IRSEP	C-15T	none	S531L	K43R	T398C
Lisboa8	BK125	G	IRS	C-15T	none	S531L	K43R	-
Lisboa4	BK674	C	IRS	C-15T	none	S531L	K43R	-
Lisboa4	BK868	L	IRSP	C-15T	none	S531L	K43R	T359C
Lisboa1	H186	H	IRSEP	C-15T	none	S531L	K43R	T2C
Lisboa1	BK616	L	IRSP	C-15T	none	S531L	K43R	392InsG
Lisboa1	BK680	L	IRSEP	C-15T	none	S531L	K43R	T359C
Lisboa1	BK536	O	IRSP	C-15T	none	S531L	K43R	250InsC

XDR-TB. In order to determine the resistance profile for second-line drugs we were able to perform drug susceptibility tests to ETH, PAS, CAP, KAN, AMK, CIP and OFX in 50 of the 58 studied isolates. We have obtained 11 different resistance patterns (Table 2.5), of which six fit into the current XDR-TB definition.[432] In total, 28 (56%) of the 50 MDR-TB isolates tested were also XDR-TB and, 32 (64%) were resistant to at least three classes of second-line drugs.

Table 2.5 - Second-line drug resistance profiles found among 50 MDR-TB isolates; CAP, capreomycin; AMK, amikacin; KAN, kanamycin; CIP, ciprofloxacin; OFX, ofloxacin; ETH, ethionamide; PAS, para-amino salicylic acid.

Resistance Pattern	No. of isolates (%)
CAP AMK ETH PAS	1(2)
CAP AMK KAN CIP OFX ETH	7(14)
CAP AMK KAN ETH	2(4)
CAP AMK KAN ETH PAS	1(2)
CAP AMK KAN OFX	1(2)
CAP CIP OFL ETH	2(4)
CAP KAN CIP OFX ETH	10(20)
CAP KAN CIP OFX ETH PAS	5(10)
CIP OFX ETH	4(8)
ETH	11(22)
KAN CIP OFX ETH	3(6)
Pan-susceptible	3(6)
Total	**50**

2.5 DISCUSSION

In the present study we have genotyped 58 MDR-TB isolates recovered in the same year, by MIRU-VNTR, and characterized the mutations associated with drug resistance in each isolate. This would serve as a starting point not only to evaluate the prevalence of the mutations associated with resistance, but also to confirm cluster integrity in order to give a more accurate vision of the transmissibility of the above strains. These 58 MDR-TB isolates, a subset of 116 MDR-TB clinical isolates, were detected by testing 1 053 isolates for drug susceptibility. This, points to a MDR-TB prevalence rate of 11% among isolates from Lisbon Health Region, which is significantly higher than the MDR-TB prevalence rate officially reported by Portuguese health authorities in the same year for the districts comprehended by this region (3.0%, 1.7% and 4.4% for Lisboa, Santarém and Setúbal districts respectively).[452] Moreover, only 17 MDR-TB cases were reported for the entire country by Portuguese health authorities in 2003, while we have isolated 116 MDR-TB isolates (each corresponding to a case) only from Lisbon health region in the same year. The MDR-TB status in Lisbon health region and in the entire country may therefore be underestimated.[453]

We have characterized the mutations associated with resistance to INH, RIF, STP and PZA. Unexpectedly the most prevalent mutation found conferring resistance to INH occurred in the promoter region of *mabA-inhA.* This is in contradiction with published studies, where the most prevalent mutation found is in *katG*.[226-229, 232, 454, 455] Mutation C-15T in *inhA*, which was detected in 53 (91.4%) isolates, has been considered a hypermorphic mutation, resulting in higher levels of InhA.[456] Overexpression of *inhA* has also been associated with INH resistance, and more recently Vilchéze *et al* have resolved the primary target of INH as being InhA.[206, 238] Nevertheless, overexpression of *inhA* was also associated with ETH resistance, an INH second-line drug analogue that shares with the former part of its action mechanism.[207, 238] The fact that the studied isolates present a high prevalence of a mutation that causes overexpression of *inhA*, may also account for the high ETH resistance prevalence (92%).

Mutations in RRDR region of *rpoB* leading to resistance to RIF were detected in 57 (98.3%) isolates, while the remaining isolate had a wild-type RRDR, which is in accordance with previous results.[250, 439] Mutation S531L was present in 53 (91.4%) isolates, being the most prevalent mutation associated with RIF resistance. Such prevalence rate is significantly higher than the prevalence rates reported in previous studies, which varied between 29% and 57%, in clinical isolates or *in vitro* generated mutants.[250, 439, 457, 458] However, the high prevalence of mutation S531L cannot be explained as a result of character ancestrality, since isolates displaying this mutation originated from several diverse clusters, not exclusively associated with MDR-TB or even any resistance pattern (data not shown) and, in the same cluster other mutations than S531L may exist.

We have screened the STP-resistant isolates included in this study, for mutations conferring resistance to STP in *rpsL* and *rrs* genes. However, we have only found one mutation (K43R in RpsL) previously associated with STP resistance.[311, 313, 323, 459] Since the remaining isolates (33.3%) did not exhibit any mutation in *rpsL* and *rrs* genes, the resistance phenotype may be due to alterations in the permeability to the drug as previously reported.[313] Another factor that could account to the absence of mutations in these STP-associated genes may be related to the critical concentration of STP used in the DST test. The critical concentrations used in this study are in accordance with those recommended by the system manufacturer's instructions. However, such concentration may not be adequate for testing the studied population, resulting in false-resistant isolates (classified as major errors).

Genetic alterations conferring PZA resistance occur mainly in the pyrazinamidase/nicotinamidase encoding gene (*pncA*), although PZA-resistant clinical isolates

with a wild-type *pncA* gene have been reported, no other genetic mechanism of resistance is currently known for PZA.[348, 350, 440, 441] All PZA-resistant isolates included in this study exhibited mutations in the *pncA* open-reading frame. Three of the frameshift mutations found have never been previously described; two were associated with medium- (282InsT; >300 mg/L) and high-level (381Del9bp; >900 mg/L) resistance. Even though a wide variety of mutations exist, several authors have reported a clustering of these mutations around three regions, codon positions 3 to17, 61 to 85 and 132 to 142.[332, 346, 350] Of the mutations obtained in the present study, only four represented by seven isolates (14.7%), are localized in these regions or in the proximity.

Although some commercial molecular methods for the detection of resistance already exist, the predominance of mutations conferring drug resistance varies geographically. The knowledge of the prevalence of resistance-associated mutations may therefore allow the development of in-house methods, locally adapted for the early detection of resistance.[292, 332, 454, 459-463] The development of a strategy that allows the early detection of the majority of the resistant isolates, in the Lisbon Health Region context, is underway.

Another application concerning the comprehension of resistance-conferring mutations in a given population is its use as secondary epidemiologic markers in the discrimination of genetic clusters, as previously reported.[405] All clinical isolates in this study have been genotyped by MIRU-VNTR and all genetic clusters obtained were also analysed considering the mutations found. Several clusters contained isolates in which sequential acquisition of resistance could be observed, namely the acquisition of resistance to EMB or PZA and subsequent spreading in the community; resistance to PZA, shown the great mutational diversity in the same cluster, generally seems to be the last acquired resistance. A cluster's genetic pool diversification may result in a possible increase in the relative fitness of such strains; Lisboa3 cluster is the most represented cluster and the one that shows the highest diversity.

Comparing with previous studies involving the characterizations of strains from Lisbon Health Region, we have determined that Lisboa3 remains as the most predominant cluster, not only in Lisboa family, but also in the context of the Lisbon Health Region super family of isolates.[26, 405] For that reason, we conclude that strains belonging to this cluster have never been contained, continuing to cause more MDR-TB cases in the community, and at this stage, probably throughout the country.

We have verified that the observed clusters could be divided into smaller clusters composed by mutational identical isolates. The combination of the MIRU-VNTR genotyping with a

mutational analysis therefore allowed the discrimination of what would be considered an outbreak or a series of possible transmission events into minor ones. Recently, a new set of mycobacterial interspersed repetitive *loci* have been proposed for standardization of MIRU-VNTR technique, this new adjustment may significantly improve the technique discrimination power *per se*.[56]

Another aspect focused on this study was the prevalence of isolates with extensive drug resistance (XDR-TB) among the studied MDR-TB isolates, which was found to be 56%. Previous studies in other countries indicated XDR-TB (at the time still considered as MDR-TB resistance plus resistance to at least three classes of second-line drugs) prevalence of 10.9%, 19%, 15% and 4% among MDR-TB isolates in Iran, Latvia, South Korea and United States respectively.[443, 464] The prevalence of XDR-TB here obtained taking in account the former definition is 64%, which is significantly higher than any XDR-TB prevalence ever reported. This high level of XDR-TB strains is probably a result of a more than 10 years of MDR-TB strains circulating in Portugal.[26] These MDR-TB strains were first identified in the beginning of the 90s, as belonging to a particular family of strains genetically related, family Lisboa. The continuous dissemination of these MDR strains in the community has permitted the progressive acquisition of resistances to second-line drugs. Another issue that might be a contribution to the high prevalence of XDR-TB is related to the high frequency of FQs prescription in other infectious diseases. Portugal is the EU country with the highest rate of FQs prescription, about 3,10 DID (diary dose per 1000 inhabitants /day).[465] During the 1990s there was a gradual increase in the general use of quinolones for a systemic use, especially second-generation quinolones, and in the current millennium, third-generation quinolones for the treatment of respiratory tract infections. Excessive use of quinolones is inevitably associated with development of resistance and may contribute to selection of resistant *M. tuberculosis* strains.

In conclusion, MDR-TB cases in Lisbon Health region are due to *M. tuberculosis* strains belonging to a family of several clusters identified several years ago as already related to MDR-TB. The non-containment of such strains will continue to cause more MDR-TB cases as these strains spread through the local community, hospitals or even throughout the country. The development of a locally adapted and rapid molecular diagnosis method is necessary for the implementation of an effective MDR-TB fast-track plan. This will allow a more rapid detection and prevent a further increase of the already high, XDR-TB rate. The present study contributes not only with the knowledge of the present situation, but also with a first step towards the creation of such strategy.

2.6 ACKNOWLEDGEMNTS

This work was supported by Fundação para a Ciência e a Tecnologia (FCT grant POCTI/ESP/42941/2001).

The authors wish to thank to the working group that participate in FCT project (POCTI/ESP/42941/2001): Hospital Pulido Valente, Instituto de Higiene e Medicina Tropical, Hospital de S. Bernardo, Hospital de St. Marta, Hospital S. João de Deus, Hospital Curry Cabral, Hospital Fernando da Fonseca, Hospital Sta. Maria, Hospital de Torres Vedras, and Laboratório de Pneumologia da Sub-região de Saúde de Lisboa, for data collection and processing; and to the personnel of Unidade de Micobactérias of Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, whose high quality of service and cooperation have made this work possible.

CHAPTER 3

Genetic Analysis of Extensively Drug-Resistant *Mycobacterium tuberculosis* strains in Lisbon, Portugal

João Perdigão[1], Rita Macedo[2], Ana Malaquias[1], Ana Ferreira[1], Laura Brum[2], and Isabel Portugal[1,2]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Unidade de Micobactérias, Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal

Journal of Antimicrobial Chemotherapy 65(2) 224-7

3.1 SYNOPSIS

Objectives: XDR-TB threatens the global control of TB worldwide. Lisbon has a high XDR-TB rate (50% of the MDR-TB), which is mainly associated with Lisboa family strains. Few studies have addressed the identification of mutations associated with resistance to second-line injectable drugs and, the relative frequency of such mutations varies geographically. The aim of this study was to characterize the genetic changes associated with the high number of XDR-TB cases in Lisbon.

Methods: In the present study we have analyzed 26 XDR-TB clinical isolates. The *gyrA*, *tlyA* and *rrs* genes were screened for mutations that could be responsible for resistance to FQs and second-line injectable drugs. Moreover, the strains under analysis were also genotyped by MIRU-VNTR.

Results: The mutational analysis identified the most frequent mutations in the resistance-associated genes: S91P in *gyrA* (42.3%); A1401G in *rrs* (30.8%) and, Ins755GT in *tlyA* (42.3%). The occurrence of mutations in *rrs* was associated with the non-occurrence of mutations in *tlyA*. The genotypic analysis revealed that the strains were highly clonal, belonging to one of two MIRU-VNTR clusters, with the largest belonging to family Lisboa. Association between mutations in *gyrA* and *rrs* or *tlyA* was verified.

Conclusions: The association of specific mutations highlighted the strains' high clonality and indicates recent XDR-TB transmission. In addition, the identification of the most frequent resistance-associated mutations will be invaluable in applying XDR-TB molecular detection tests in the region in the near future.

3.2 INTRODUCTION

XDR poses a serious threat to TB management. XDR-TB is defined as multidrug resistance MDR, *i.e.* resistance to at least INH and RIF, plus resistance to a FQ and any of the injectable second-line drugs (AMK, CAP or KAN).[466]

At least four different antimycobacterial drug classes are required for an adequate treatment and in the majority of the XDR-TB cases there are not enough options.[467] Resistance to FQs and second-line injectable drugs is indeed associated with poor outcomes.[468] As a consequence, such strains are not only very difficult to treat but also to eliminate.

In recent years, Portugal has had TB rates above the EU average, especially if we consider only the first 15 countries that have joined the EU (EU-15) and in which Portugal is included.[469]

Ten years ago, some of us described a family of strains causing a MDR-TB outbreak in the Lisbon Health Region. This family was later designated as the Lisboa Family. [26]

More recently, in one of our latest reports, we analysed a panel of multidrug-resistant isolates and observed that ≈ 50% were XDR-TB isolates; to our knowledge the highest XDR-TB rate ever reported.[24] However, the genetic basis of such resistance was unknown.

In this study, and given the paucity of studies characterizing such mutations in different settings, we aimed at determining which genetic changes are responsible for resistance to injectable second-line drugs and FQs and, association with specific clusters in Lisbon, Portugal.

3.3 METHODS

Clinical Isolates

We have analyzed 26 XDR-TB clinical isolates, each corresponding to a different patient. These isolates were recovered from several hospital units and laboratories of Lisbon's Health Region during the year of 2005, and correspond to 74.3% of all the XDR isolated in this region, in this year. The remaining isolates were unviable and so were not processed. Our laboratory received 67 new and previously treated MDR-TB cases, 35 (52.2%) being XDR-TB.

DST. Isolates were tested for first-line drug susceptibility by the BACTEC 960 MGIT methodology according to the manufacturer's instructions. MDR-TB isolates were tested for second-line drug susceptibility by the radiometric BACTEC 460, as previously described.[24]

PCR Amplification and DNA Sequencing

A 981 bp fragment, containing the *tlyA* gene, was amplified by PCR using *tlyA*F3 (AAGGCATCGCACGTCGTCTTTCC) and *tlyA*R3 (TGTCGCCCAATACTTTTTCTACGC) primers; the region comprehended between the positions 1151 and 1590 of the *rrs* gene was amplified by PCR using primers RRS2-F (TGCCGGGGTCAACTCGGAGG) and RRS2-R (GAACCCCTCACGGCCTACGC); and, a 320 bp (78-397) internal fragment of the *gyrA* gene was amplified by PCR with oligonucleotide primers *gyrA*-F (CAGCTACATCGACTATGCGA) and *gyrA*-R (GGGCTTCGGTGTACCTCAT). Cycling conditions were as follows: denaturation at 94ºC for 4 minutes; 35 cycles of denaturation at 94ºC for 1 min, annealing at 58ºC (55ºC for *gyrA*) for 1 min and extension at 72ºC for 2 min and 30 sec (1 min for *gyrA*); and, lastly, a final extension step at 72ºC for 10 min was performed. Crude DNA extracts of each isolate were used in all molecular biology reactions.

Prior to sequencing, the amplified fragments were purified using the MSB™ Spin PCRapace (Invitek™). The sequencing reactions were carried out with BigDye™ Terminator Cycle Sequencing Kit with AmpliTaq DNA polymerase (Applied Biosystems™) using primers *tlyA*F1 (GTTGTTGGCCGCCCTGGAGT) and *tlyA*R1 (GGTCTCGGTGGCTTCGTCGC), RRS2-R, or *gyrA*-F as appropriate.

All the mutations found were compared with those included in the Tuberculosis Drug Resistance Mutation Database (www.tbdreamdb.com).

MIRU-VNTR

All isolates were also genotyped by the 12 *loci* Mycobacterial Interspersed Repetitive Unit – Variable Number of Tandem Repeats (MIRU-VNTR) as described previously.[24]

Statistical Analysis

The statistical association between clusters and/or mutations was done using a two-tailed Fisher's Exact Test in GraphPad Prism™ Version 4.00 (GraphPad Software, Inc).

3.4 RESULTS

For this study we analysed 26 different XDR-TB isolates collected from Lisbon Health Region's hospital units during 2005. In order to determine the possible molecular basis for resistance to FQs and injectable second-line drugs, the isolates were screened for mutations in three genes – *gyrA*, *rrs* and *tlyA*. We preferentially screened the *gyrA* gene and not *gyrB* in this set of isolates since previous results on Lisbon FQ-resistant isolates indicates that *gyrB* mutations associated with resistance are absent in these strains (R. Macedo, unpublished results). We identified three different missense mutations in GyrA (D94A, D94G and S91P), the most common being mutation S91P which was present in 11 (42.3%) isolates (Table 3.1).

Concerning the mutations associated with resistance to injectable second-line drugs, all the analyzed isolates, except one, had mutations occurring in either *tlyA* or *rrs* genes. Occurrence of a mutation in *tlyA* gene was associated with the non-occurrence of a mutation in *rrs* gene (Fisher´s Exact Test, p=0.0010). The most prevalent mutations found in those genes were A1401G in *rrs* and a newly described frameshift mutation in *tlyA*, Ins755GT, which alone was present in 42.3% of the analysed isolates (Table 3.1).

When comparing genotypes, isolates were divided into two large MIRU-VNTR clusters (Table 3.2). One of the clusters belongs to the highly predominant Lisboa family and the other was previously designated by Q1 and presents a high genetic similarity with Lisboa strains. We verified that mutations occurring in *gyrA*, *rrs* and *tlyA* genes were cluster-associated. GyrA D94A and *rrs* A1401G occurred exclusively in cluster Q1, and GyrA D94G, S91P and *tlyA* Ins755GT were exclusive to the Lisboa3 cluster (Table 3.2). Different mutations at the same codon (GyrA94) were concordant with the MIRU-VNTR clustering analysis. Eleven of 17 FQ-resistant isolates on Lisboa3 cluster had the GyrA S91P mutation, which was strongly associated with *tlyA* Ins755GT mutation (Fisher´s Exact Test, p<0.0001). In cluster Q1, eight isolates had the GyrA D94A mutation, which was associated with *rrs* A1401G mutation (Fisher´s Exact Test, p<0.0001). Such isolates, in both clusters, may therefore illustrate several cases of primary XDR-TB transmission, although no patient data was available to establish epidemiological relatedness or provide evidence regarding the chain of transmission.

None of the mutations associated with resistance found in this study were detected in susceptible strains.

Table 3.1 – Mutational frequency found on the resistance-associated genes of the 26 studied isolates.

Gene	Mutation (nt)	Mutation (aa)	No. of Isolates (%)
gyrA			
	T271C	S91P	11 (42.3)
	A281C	D94A	8 (30.8)
	A281G	D94G	6 (23.1)
	None	None	1 (3.8)
		Total	**26 (100)**
rrs			
	G1158T	-	1(3.8)
	A1401G	-	7(26.9)
	A1401G+G1158T	-	1(3.8)
	G1484T	-	1 (3.8)
	G1484T+C1402A	-	1 (3.8)
	Ins1211G	-	1 (3.8)
	Ins1512G	-	1 (3.8)
	None	-	13 (50.0)
		Total	**26 (100)**
tlyA			
	A205C	K69Q	1 (3.8)
	G223T	E75X	2 (7.7)
	Ins521CG	Frameshift	1 (3.8)
	Ins755GT	Frameshift	11 (42.3)
	None	None	11 (42.3)
		Total	**26 (100)**

Table 3.2 – Resistance patterns, clusters and mutations found in the analysed isolates.

Isolate	First-Line Resistance Profile[a]	Second-Line Resistance Profile[b]	Mutations[c]					MIRU-VNTR Cluster
			inhA	*rpoB*[d]	*gyrA*	*tlyA*	*rrs*	
BK519	IRSEP	CAP, KAN, CIP, OFX, ETH,	C-15T	C1349T (S531L)	A281G (D94G)	none	C1402A; G1484T	Lisboa3
BK128	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281G (D94G)	G223T (E75X)	G1158T	Lisboa3
BK87	IRS	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281G (D94G)	none	G1484T	Lisboa3
BK530	IRSEP	CAP, CIP, OFX, ETH	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	Ins1211G	Lisboa3
BK872	IRSEP	CAP, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	Ins1512G	Lisboa3
BK103	IRSEP	CAP, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281G (D94G)	A205C (K69Q)	none	Lisboa3
BK232	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH, PAS, CS	C-15T	C1349T (S531L)	A281G (D94G)	G223T (E75X)	none	Lisboa3
BK643	IRSEP	CAP, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281G (D94G)	Ins521CG	none	Lisboa3
BK30	IRSEP	CAP, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK81	IRSEP	CAP, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK127	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH, CS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK214	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH, PAS, CS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK309	IRSEP	CAP, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK548	IRS	CAP, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK801	IRSE	CAP, CIP, OFX, ETH	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK839	IRSEP	CAP, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK1144	IRSEP	CAP, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	T271C (S91P)	Ins755GT	none	Lisboa3
BK104	IREP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK269	IREP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	A1304T (D516V)	A281C (D94A)	none	A1401G	Q1
BK364	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH, PAS, CS	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK460	IREP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK474	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK667	IREP	CAP, AMK, KAN, CIP, OFX, ETH, PAS	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK733	IREP	CAP, AMK, KAN, CIP, OFX, ETH	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G	Q1
BK396	IRSEP	CAP, AMK, KAN, CIP, OFX, ETH,	C-15T	C1349T (S531L)	A281C (D94A)	none	A1401G; G1158T	Q1
BK372	IR	CAP, AMK, KAN, CIP, OFX, ETH,	C-15T	C1349T (S531L)	none	none	none	Q1

[a] I – isoniazid; R – rifampicin; E – ethambutol; S – streptomycin; P – pyrazinamide.

[b] CAP – capreomycin; AMK – amikacin; KAN – kanamycin; CIP – ciprofloxacin; OFX – ofloxacin; ETH- ethionamide; PAS – para-amino salicylic acid; CS – cycloserine.
[c] Nucleotide mutation with resulting amino acid mutation in brackets, if applicable.
[d] nucleotidic mutational position relative to *M. tuberculosis* H37Rv *rpoB*; aminoacidic positioning is relative to *E. coli* numbering.

3.5 DISCUSSION

We have examined the main genes associated with resistance to FQs and injectable second-line drugs – *gyrA*, *rrs* and *tlyA*, respectively. The mutations found in the *gyrA* gene have been previously described as associated with FQ resistance, and our results support this.

In contrast, the screening of *rrs* and *tlyA* yielded mutations not previously described. Twenty-five isolates displayed mutations in one of these genes and only three isolates had simultaneous mutations in both genes. Our results also indicate that only a subset of the mutations conferring KAN-resistance actually confers resistance to AMK.[359, 470] The data seem to support the perception that resistance to AMK is associated with A1401G mutations but not with *tlyA* mutations, while nine KAN-resistant isolates exhibited *tlyA* mutations only. However, in a study by Maus *et al* nonsense *tlyA* mutants remained susceptible to KAN and AMK, which indicates that *tlyA* mutations probably have no role in resistance to these second-line aminoglycosides.[375] We cannot, however, exclude the existence of mutations outside the analyzed *loci*, since mutants resistant to KAN with no *tlyA* or *rrs* mutations have been found. In addition, our results support previous findings in that mutations in *rrs* can mediate cross-resistance between CAP, KAN and AMK.[359] On the other hand, our results clearly differ from the results obtained by Jugheli *et al* who have found no mutations in the *tlyA* gene of 145 isolates of which 78 and 66 were KAN- and CAP-resistant, respectively.[470]

A new reverse hybridization based test to screen for cultured XDR-TB isolates - GenoType MTBDR*sl* (Hain Lifescience, GmbH, Nehren, Germany)[471] is already available commercially. This new test detects mutations associated with resistance to second-line injectable drugs and FQs. According to results obtained in our study, this test would have been able to detect 96.2% of the FQ resistance and 34.6% of the resistance to CAP, KAN or AMK. Since resistance-associated mutations vary geographically we therefore stress the need for knowledge concerning the prevalence of such mutations before *in situ* implementation of these commercial assays.

Another important conclusion concerns the clonality of the analysed strains. These strains were shown to be highly clonal, belonging to one of two close genetic clusters. Also important,

was the evidence that a large proportion of the analysed cases (73.1%) were possibly primary XDR-TB. This shows that the appropriate means to avoid transmission are not in place since active transmission of these deadly forms of TB is taking place.

The strains analysed in this study are highly resistant, presenting clinicians with few options. Four to six drugs would be necessary to treat these patients and only 4 of the 26 isolates in this study appear to be susceptible to at least four of these drugs. In the remaining isolates, the so-called third-line drugs (CFZ, linezolid, amoxicillin/clavulanate, imipenem, macrolides), which have *in vitro* activity but with limited clinical data, should be explored.[467]

The present study has, consequently, important implications for the treatment, diagnosis and epidemiology of XDR-TB in the Portuguese setting. Also, it contributes to the global awareness of XDR/MDR-TB, and the urgency for new therapeutic options and novel molecular diagnosis tests adapted to specific settings. This study also highlights the utmost importance of strict regulation in the prescription of antibacillary drugs, without which is impossible to effectively control and eliminate TB, even with new antibacillary drugs.

3.6 ACKNOWLEDGEMENTS

The authors wish to thank to the laboratory personnel of Hospital Pulido Valente, Instituto de Higiene e Medicina Tropical, Hospital de S. Bernardo, Hospital de St. Marta, Hospital S. João de Deus, Hospital Curry Cabral, Hospital Fernando da Fonseca, Hospital Sta. Maria, Hospital de Torres Vedras, and Laboratório de Pneumologia da Sub-região de Saúde de Lisboa, for data collection and processing; and to the staff of Unidade de Micobactérias of Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, whose high quality of service and cooperation have made this work possible.

3.7 FUNDING

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT grant PTDC/BIA-MIC/71280/2006).

3.8 TRANSPARENCY DECLARATIONS

None to declare

CHAPTER 4

Tuberculosis Drug-Resistance in Lisbon, Portugal: a 6-Year Overview

João Perdigão[1], Rita Macedo[2], Carla Silva[1], Claúdia Pinto[1], Cristina Furtado[2], Laura Brum[2], and Isabel Portugal[1,2]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Departamento de Doenças Infecciosas, Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal

Clinical Microbiology and Infection 17(9) 1397-402

4.1 ABSTRACT

MDR and extensive XDR pose a serious threat to TB management in Portugal. The country has high TB incidence rates in comparison with the other European Union countries, with Lisbon Health Region being one of the most affected regions.

In the present study we have analyzed a convenience sample of 3 025 *M. tuberculosis* clinical isolates, recovered over a six-year period (2001-2006) in Lisbon Health Region, regarding drug-resistance to both first-line and second-line drugs. Moreover, one hundred of these isolates were also genotyped by 12-*loci* Mycobacterial Interspersed Repetitive Unit – Variable Number of Tandem Repeats (MIRU-VNTR) analysis.

We have compared each year and observed the existence of 22 different resistance profiles with MDR-TB rates ranging between 9.9% and 15.2% and XDR-TB rates, relative to the number of MDR-TB isolates, between 44.3% and 66.1% (excluding one year here considered as an outlier). A steady increase in the fraction of MDR-TB isolates resistant to all first-line drugs was also noticed. The genotyping analysis of MDR-TB isolates revealed six clusters, of which three (Lisboa3, Lisboa4 and Q1) were related with XDR-TB.

Our results show that active transmission of MDR- and XDR-TB is taking place and that the high prevalence of observed XDR-TB is due to the continued transmission of particular genetic clusters. An early detection of resistant cases would greatly benefit from the enforcement in case detection and implementation of genotyping in diagnostic routines.

4.2 INTRODUCTION

The worldwide control of TB infection and transmission may be severely hampered by the recurrent development of resistance to antibacillary drugs. MDR and XDR constitute a special and more serious case of resistance because it renders ineffective the most efficient first- and second-line drug therapies, respectively.

Although TB cases resistant to second-line drugs were widely known, the growing awareness concerning the widespread and growing rates in industrialized nations highlighted the necessity to recognize and classify this old problem.[443] According to the latest WHO report, at least one XDR-TB case has been described in 58 countries.[472] Other reports also highlight XDR-TB as a very serious global public health crisis. The survival rates associated with XDR-TB are low compared with MDR-TB or susceptible TB. Population studies performed in Russia, South Korea and Peru indicated survival rates of 48%, 54% and 60%, respectively among HIV-negative patients.[473-475]

The tuberculosis situation in Portugal is worrisome. The incidence rate is high, taking in account the incidence rates of the other European Union members and the recent reports by some of us concerning the high prevalence of MDR-TB and XDR-TB in the Lisbon Health Region.

Genotyping of clinical isolates would allow a better identification of transmission chains. The main genetic clusters present in the region were already identified and, the majority described more than ten years ago. Such isolates apparently continue to circulate in the region and have inclusively spread to other regions in the country without noticeable limitations.[24, 26]

The INSA-RJ laboratory has, over the years, collected and analyzed a high number of TB clinical isolates, mainly from Lisbon Health Region. In the present report, we analyze the prevalence of resistant isolates, including MDR- and XDR-TB, over a six-year period (2001-2006).

4.3 METHODS

Clinical Isolates

We have analyzed the data records of all clinical isolates sent for identification and drug susceptibility tests at the INSA-RJ, Portugal's national TB Reference Laboratory. For the entire study, a convenience sample of 3 025 isolates were selected from our tuberculosis database.

The selection of these isolates was based on: complete information concerning drug susceptibility testing; isolation in Lisbon Health Region. Isolates with no *Mycobacterium tuberculosis* complex identification were discarded and, duplicate isolates were removed from the data set. These criteria allowed the selection of 283 isolates from 2001, 741 from 2002, 667 from 2003, 547 from 2004, 442 from 2005 and 345 from 2006.

DST

All isolates were tested for susceptibility to first-line drugs (INH, RIF, STP, EMB and PZA) by the BACTEC 960™ MGIT™ or BACTEC 460™ (Becton-Dickinson™) method according to the manufacturer's indications. Isolates resistant to INH and RIF were considered MDR-TB and tested for second-line drug susceptibility.

Second-line DST was performed by the BACTEC 460™ system to CAP (1,25 mg/L), AMK (1,0 mg/L), KAN (5,0 mg/L), CIP (1,0 mg/L), OFX (2,0 mg/L), ETH (5,0 mg/L) and PAS (4,0 mg/L), according to Pfyffer *et al.* [446] MDR-TB isolates resistant to, at least, a FQ (OFX or CIP) and a second-line injectable drug (CAP, KAN or AMK) were considered XDR-TB isolates.

MIRU-VNTR Genotyping

One hundred isolates (28 from 2004, 48 from 2005 and, 24 from 2006) were genotyped by the automated 12-*loci* MIRU-VNTR technique as described by Supply *et al.*[57]

4.4 RESULTS

For this study, the records of a total of 3 025 clinical isolates were analysed. The analysis of these isolates allows us to assess the TB situation regarding drug resistance among all isolates tested at the TB National Reference Laboratory from Portugal's National Institute of Health Dr. Ricardo Jorge. This convenience sample may also provide a general picture of the prevalence of MDR-TB and XDR-TB in Lisbon Health Region.

We started by analyzing the isolates resistant to first-line antibacillary drugs. Overall, we found 22 different resistant profiles in isolates resistant to one or more first-line drugs distributed through the years (2001-2006). Eight profiles met the criteria for MDR-TB, with 2003 registering the highest number of MDR-TB cases (82) (Table 4.1).

Table 4.1 – First-line resistant profiles found among the analyzed isolates in each year.

Resistance Profile[a]	No. of clinical isolates studied per year					
	2001	**2002**	**2003**	**2004**	**2005**	**2006**
Pan-susceptible	220	548	503	424	331	268
E	0	0	0	0	0	2
ES	0	0	0	0	0	1
I	4	26	6	12	14	5
ISE	0	0	0	0	1	0
ISEP	0	0	0	2	0	0
IP	2	2	1	2	1	1
IR	2	10	3	0	3	3
IRE	0	0	0	0	1	1
IREP	3	4	0	2	9	3
IRSE	4	2	6	4	2	0
IRSEP	11	21	32	24	30	18
IRP	1	7	7	4	4	1
IRS	7	6	7	4	5	1
IRSP	12	31	27	19	13	7
IS	5	13	16	16	9	10
ISP	0	2	1	0	1	0
P	0	0	0	0	5	0
R	0	3	2	3	0	1
RP	0	0	1	0	0	0
RS	0	1	0	0	0	0
S	12	65	54	31	12	23
SP	0	0	1	0	1	0
Total	283	741	667	547	442	345

[a] - I – isoniazid, R – rifampicin, S – streptomycin, E – ethambutol and P – pyrazinamide.

The analysis of the susceptibility to second-line drugs yielded 26 different resistant patterns, 15 of which corresponded to XDR-TB isolates (Table 4.2). From 2001 to 2006, 14 (2006) to 47 (2003) XDR-TB cases were detected in this study.

Table 4.2 – Second-line resistant profiles found among the analyzed isolates in each year.

Resistance Profiles [a]	No. of clinical isolates studied per year					
	2001	2002	2003	2004	2005	2006
Pan-susceptible[a]	2	6	7	7	12	3
CAP AMK CIP OFX ETH	0	0	0	1	0	1
CAP AMK ETH PAS	0	0	1	0	0	0
CAP AMK KAN CIP ETH	1	1	0	0	0	0
CAP AMK KAN CIP ETH PAS	0	3	0	0	0	0
CAP AMK KAN CIP OFX ETH	4	5	10	14	11	3
CAP AMK KAN CIP OFX ETH PAS	0	0	1	5	7	2
CAP AMK KAN CIP OFX PAS	0	0	0	0	1	1
CAP AMK KAN ETH	0	1	4	1	2	0
CAP AMK KAN ETH PAS	0	0	1	0	0	0
CAP AMK KAN OFX	0	0	1	0	0	0
CAP CIP ETH	1	0	1	0	0	0
CAP CIP ETH PAS	0	1	0	0	0	0
CAP CIP OFX ETH	2	3	3	1	1	0
CAP ETH	0	0	0	0	1	0
CAP ETH PAS	0	0	0	0	1	0
CAP KAN CIP OFX ETH	4	6	20	8	7	3
CAP KAN CIP OFX ETH PAS	0	5	7	1	6	2
CAP OFX ETH	0	0	0	1	0	0
CIP ETH	0	2	1	1	1	0
CIP OFX	0	1	0	1	0	0
CIP OFX ETH	5	4	5	2	4	1
CIP OFX ETH PAS	0	4	1	0	0	0
ETH	9	35	17	10	17	2
KAN CIP OFX ETH	6	6	4	5	2	1
KAN CIP OFX ETH PAS	0	5	1	1	0	1
ND	249	653	581	488	369	325
OFX	0	0	1	0	0	0
Total	283	741	667	547	442	345

[a] CAP – capreomycin; AMK – amikacin; KAN – kanamycin; CIP – ciprofloxacin; OFX – ofloxacin; ETH- ethionamide; PAS – para-amino salicylic acid; ND – not determined.

The trend in the MDR-TB and XDR-TB cannot be clearly evaluated since the varying number of clinical samples received by the laboratory for testing may act as a confounding factor in such analysis. Nevertheless, considering the years of 2002-2005, an increase in the MDR-TB may have occurred because in 2005 the MDR-TB cases accounted for 15.3%, an increase of approximately 4% considering the three previous years (Table 4.3). Also, the percentage of MDR-TB isolates with resistance to all first-line drugs appears to show a steady increase from 2001/2002 to 2006 varying between 25.9% in 2002 and 52.9% in 2006. Considering the XDR-TB rate, and taking into account the year of 2006 as an outlier given the low number of tested strains, we verify that it varied between 44.3% (2002) and 66.1% (2004) with a mean value of 54.3%.

Twelve-*loci* MIRU-VNTR genotyping analysis was implemented in 2004, and for that reason it was not possible to assess the genetic clusters circulating in the previous years. Genotyping data was available for one hundred of the selected MDR-TB isolates from 2004-2006. A MIRU-VNTR dendrogram of the genotyped isolates was produced for each year. Genetic clusters were defined as a group of two or more isolates exhibiting the same MIRU-VNTR pattern. A clear predominance of Lisboa family strains was verified between 2004 and 2006. Six clusters were identified, with three belonging to Lisboa family (Lisboa3, Lisboa4 and Lisboa10) (Figure 4.1). Lisboa3, Lisboa 4 and Q1 clusters were associated with XDR-TB, with Lisboa3 having the highest number of XDR-TB cases (Table 4.4).

Table 4.3 – Number and percentage of resistant, MDR- and XDR-TB isolates, distributed by each year (2001-2006).

Resistance	No. of isolates (%)					
	2001	**2002**	**2003**	**2004**	**2005**	**2006**
Total	283	741	667	547	442	345
Resistant[a]	63 (22.3)	193 (26.0)	164 (24,6)	123 (22.5)	111 (25.1)	77 (22.3)
MDR-TB[b]	40 (14.1)	81 (10.9)	82 (12.3)	57 (10.4)	67 (15.3)	34 (9.9)
XDR-TB[c]	18 (52.9)	35 (44.3)	47 (57.3)	37 (66.1)	34 (50.7)	14 (73.6)

a – Isolates resistant to one or more antibacillary drugs;
b – Isolates resistant to at least INH and RIF;
c – XDR-TB rate expressed as the fraction of the tested MDR-TB isolates with resistance to at least one second-line injectable drug and a FQ.

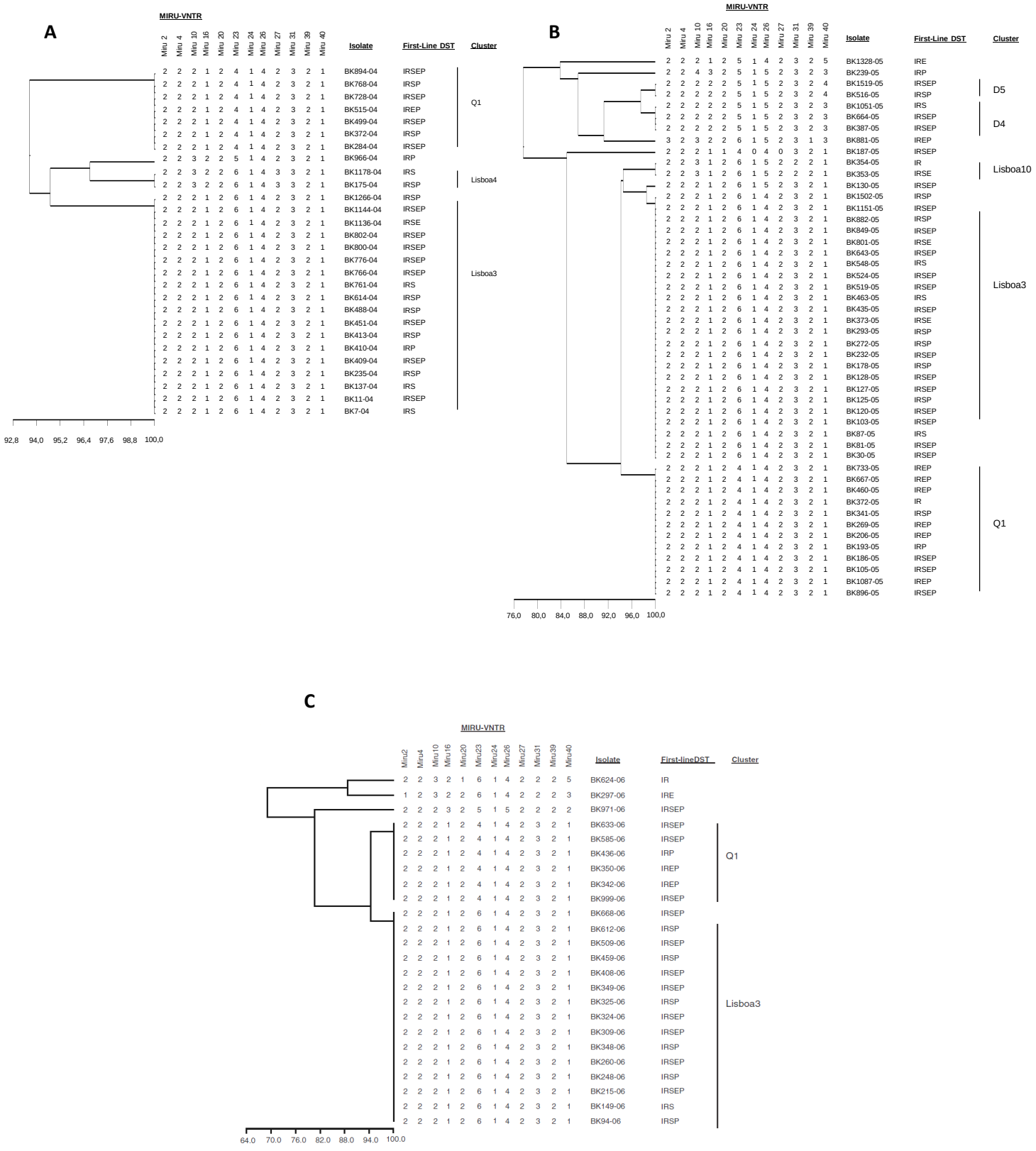


Figure 4.1 – MIRU-VNTR dendrograms of MDR-TB isolates recovered in the years of 2004 (A), 2005 (B) and, 2006 (C). First-Line DST: I – isoniazid, R – rifampicin, S – streptomycin, E – ethambutol and P – pyrazinamide.

Table 4.4 – Distribution of the MDR-TB genotyped isolates *per* cluster *per* year and, number of XDR-TB isolates in each cluster.

MIRU-VNTR Cluster	2004		2005		2006	
	No. of MDR Isolates	No. of XDR Isolates	No. of MDR Isolates	No. of XDR Isolates	No. of MDR Isolates	No. of XDR Isolates
D5	-	-	2	0	-	-
D4	-	-	3	0	-	-
Lisboa3	18	11	24	18	15	13
Lisboa4	2	1	-	-	-	-
Lisboa10	-	-	2	0	-	-
Q1	7	3	12	8	6	3

- Not detected in that year

4.5 DISCUSSION

Portugal is a country with a considerable TB incidence taking in account the other EU countries. The high number of XDR-TB cases and rates in an area of 14 572 km^2 (Lisbon, Setúbal and Santarém districts) are a concern to those who have to manage this problematic situation.

In a previous publication we have pointed to two possible explanations for the high prevalence of XDR-TB in Portugal: the non-containment of Lisboa Family strains and the high FQ prescription rate.[24, 465, 476] Additionally, the use of the WHO-recommended standard treatment regimen for MDR-TB suspected cases with unavailable first-line drug susceptibility tests results may potentiate this situation.[203] This treatment regimen includes a six month initial phase with KAN, ETH, a FQ, PZA and EMB. In a previous study, we have shown that 92% of the MDR-TB isolates were also resistant to ETH, mostly due to cross-resistance between INH and ETH and, that 79.4% of the same MDR-TB isolates were also resistant to EMB and/or PZA.[24] In the present study, that proportion ranged from 77.5 to 93.0% and resistance to ETH among all tested isolates was seen to range from 80.0 to 94.1%. Consequently, the use of the MDR-TB standard treatment regimen in the Portuguese setting does not provide the patient with the necessary active drugs and, may lead to resistance acquisition to second-line injectable drugs and/or FQs and hence, to XDR-TB development.

The results presented here show that, in accordance with previous studies by some of us, XDR-TB is not only a current problem that must necessarily be addressed, but is the result of a continued circulation of MDR-TB clones detected in the 1990s.[26] The main clusters were detected and shown to maintain their prevalence, over the three year-period (where genotyping was available) in the MDR-TB/XDR-TB scenario of Lisbon Health Region.

The largest cluster of XDR-TB in the world, in South Africa, belongs to the KZN family of TB strains, described in 1996 and at the time pan-susceptible or resistant to first-line drugs only.[477] In Portugal, the situation is similar, Lisboa family, which was identified in 1996, was already strongly associated with MDR-TB at that period and now is associated with XDR-TB.[24, 26, 478] Monitoring of these strains through timely genotypic analysis would provide a mean to identify more rapidly MDR-TB and XDR-TB cases and ensure greater tracing of contacts, with the aim of containing these strains, thereby avoiding a possible XDR-TB outbreak. Given the high clonality that was found associated with MDR-TB and XDR-TB, and because genotyping results can be obtained earlier than DST results, the identification of an isolate belonging to a MDR-TB or XDR-TB associated cluster might be useful in identifying a possible MDR-TB or XDR-TB strain at an earlier stage.

Transmission can also be enhanced by another factor, the sputum smear-positive case rate. Sputum smear-positivity is associated with higher transmission and Portugal has a high sputum smear-positive case rate when compared with the majority of the EU countries, which *per se* may be predictive of a worse outcome.[469, 479, 480] The explanation to higher bacillary loads might be related with the strain itself or may be due to a delayed case detection. If effective measures are not put in place to avoid the transmission of these strains, it is conceivable that an increase of the inherently high treatment costs of MDR-TB and XDR-TB for national tuberculosis control programs may occur, diverting resources necessary to control TB.[481]

A limitation to this study is the fact that a convenience sample was used. Nevertheless, the high number of cases found within such reduced geographic region justifies particular attention and urgent implementation of control measures.

A study concerning XDR-TB treatment in Tomsk, Russia by Keshavjee *et al* showed that although there was a less favorable outcome it is possible to treat XDR-TB with regimens including five active drugs plus a FQ and CAP even when resistance was found.[473] In the Portuguese setting, as we have shown, given the higher degree of resistance, the problem resides in finding at least four active drugs. This study is an important contribution for the

knowledge concerning drug resistance in Portugal and improvement of standard treatment regimens.

4.6 ACKNOWLEDGEMENTS

This work was partially supported by Fundação para a Ciência e Tecnologia (FCT grant PTDC/BIA-MIC/71280/2006) and the XDRTB Early Detection in Greater Lisbon (2009) project from the Calouste Gulbenkian Foundation. The authors wish to thank the personnel of Hospital Pulido Valente, Instituto de Higiene e Medicina Tropical, Hospital de S. Bernerdo, Hospital de Santa Marta, Hospital S. João de Deus, Hospital de Curry Cabral, Hospital Fernando da Fonseca, Hospital de Santa Maria, Hospital de Torres Vedras and Laboratório de Pneumologia da Sub-região de saúde de Lisboa, for data collection and processing, and the staff of Unidade de Micobactérias of Centro de Bacteriologia, instituto Nacional de Saúde Dr. Ricardo Jorge, whose high quality of servisse and cooperation have made this work possible.

4.7 TRANSPARENCY DECLARATIONS

The authors declare that they have no conflicts of interest in relation to this work.

CHAPTER 5

From Multidrug-Resistant Tuberculosis to Extensively Drug Resistance in Lisbon, Portugal: the stepwise mode of resistance acquisition

João Perdigão[1], Rita Macedo[1,2], Carla Silva[1], Diana Machado[3], Isabel Couto[3,4], Miguel Viveiros[3], Luisa Jordao[5] and Isabel Portugal[1]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Public Health Laboratory: Mycobacteriology/Tuberculosis, Public Health Department, Administração Regional de Saúde de Lisboa e Vale do Tejo, I.P., Lisboa, Portugal; [3]Grupo de Micobactérias, Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (IHMT/UNL), Lisboa, Portugal; [4]Centro de Recursos Microbiológicos (CREM), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; [5]Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisboa, Portugal

Journal of Antimicrobial Chemotherapy 68(1) 27-33

5.1 SYNOPSIS

The development and transmission of XDR-TB constitutes a serious threat to the effective control of tuberculosis in several countries. Resistance is the result of adaptive selection of *M. tuberculosis* strains that acquire and accumulate specific mutations at specific genes. Current mechanisms of drug resistance include the modification or overexpression of drug targets, drug efflux, inactivation of activator enzymes and overexpression of drug-modifying enzymes.

In the present study a set of 44 MDR or XDR-TB isolates were genotyped and screened for mutations in genes associated to second-line drug resistance, namely *tlyA*, *gyrA*, *rrs* and *eis* genes. We have determined the most prevalent mutations found in each gene to be Ins755GT in *tlyA*, A1401G in *rrs*, G-10A in *eis* and S91P in *gyrA*. Additionally, two genetic clusters were found in this study: Lisboa3 and Q1. The characteristic mutational profile found among recent XDR-TB circulating in Lisbon was also found in MDR-TB strains isolated in the 90s, highlighting the ancient XDR-TB problem in this region. Also investigated was the resistance level conferred by *eis* G-10A mutations, revealing that *eis* G-10A mutations may result in undetectable AMK resistance.

In conclusion, the analysis of the distribution of the mutations found by genetic cluster showed that in Q1 cluster two mutations, *gyrA* D94A and *rrs* A1401G, were enough to ensure development of XDR-TB from a MDR strain. Moreover, in Lisboa3 cluster it was possible to elaborate a model in which the development of KAN low-level resistance was at the origin of the emergence of XDR-TB strains that can be discriminated by *tlyA* mutations.

5.2 INTRODUCTION

XDR-TB is a consequence of multiple acquisitions of mutations in specific resistance-associated genes. Such mutations may interfere with drug-target interaction or, overexpress the protein target or other drug modification enzymes. Since XDR-TB is defined as resistance to INH and RIF plus resistance to a FQ and a second-line injectable antibacillary drug - CAP, AMK or KAN, the most frequently implicated genes are: *inhA*, *katG*, *rpoB*, *gyrA*, *rrs*, *tlyA* and *eis*.[359, 375, 432]

According to the latest drug resistance surveillance report from the WHO, XDR-TB cases have already been reported in over 58 countries. The worldwide occurrence of XDR-TB raises serious public health concerns as this level of cumulative resistance to the most effective and available antibacillary drugs is most often found to be untreatable.[472]

In Portugal, the high XDR-TB rates in its capital city Lisbon, also the region with the highest TB incidence, have shown to be mainly associated with particular genetic clusters that have underwent extensive clonal expansion in the community. Of special importance are the strains belonging to some of the Lisboa family genetic clusters that are in circulation for almost two decades. XDR-TB in the region has been considered a serious public health concern in the years following the first definition of XDR-TB, but recent studies now show that this is an old problem that can be, at least, tracked back to the 90s.[26, 482]

The main genetic clusters associated with MDR and XDR-TB in Lisbon, designated Lisboa3 and Q1, have been described in previous publications. Lisboa3 cluster was originally described in the 90s, highly associated with the original MDR-TB outbreak in the region, and presently with XDR-TB. The Q1 cluster, detected only in strains recovered from 2003, is genetically very close to the original Lisboa family and contains XDR-TB strains, although presenting a lesser mutational diversity when compared with Lisboa3 cluster.[24-26, 405] The exact phylogenetic origin of this later cluster in respect to Lisboa family is still under study as there are significant genetic differences regarding to other genomic *loci* or spoligotyping profiles (unpublished data).

Previous data from our laboratory has contributed to elucidate the molecular basis of resistance development among MDR-TB strains and XDR-TB in the region through the mutational analysis of *gyrA*, *rrs* and *tlyA*.[25] However, the mode of resistance acquisition and the genetic determinants leading to second-line aminoglycoside resistance in the region are not fully understood. Combining both mutational and genotypic approaches will eventually

lead to insights on how resistance develops and allows further strain discrimination for the study of transmission in the community.[405]

Zaunbrecher *et al*, associated promoter mutations in *eis* with KAN resistance. The gene product of *eis*, an acetyl-transferase, catalyzes multiacetilation of aminoglycosides, modifying KAN and, with lesser affinity, AMK. The explanation offered by the authors suggested that *eis* promoter mutations are responsible for KAN resistant isolates that do not present cross-resistance to AMK.[371, 372]

Here, in attempt to further elucidate the dynamics of resistance acquisition and investigate an eventual role for *eis* promoter mutations in second-line aminoglycoside resistance we have studied a set of MDR/XDR-TB isolates that included some of the original strains involved in the Lisbon MDR-TB outbreak from the 90s.[26]

5.3 METHODS

Clinical Isolates, Drug Susceptibility Testing and Genotyping

A total of 44 MDR/XDR *M. tuberculosis* clinical isolates recovered from TB patients of the Lisbon region, between 1997 and 2009, were studied. All strains were isolated at local hospital units and laboratories and sent for drug-susceptibility testing at Level 3 Mycobacteriology Laboratories. For strains isolated in 1997-1998 and in 2005, only DNA was available for mutational analysis and genotyping.

At the time of isolation, first-line DST for isolates recovered between 1997-98 was performed using the radiometric BACTEC™ 460 system (Becton Dickinson Diagnostic Systems, Sparks, MD, USA) for INH, RIF, EMB and STP. For strains isolated after 2005, first-line DST for the remaining isolates was carried out by the fluorimetric BACTEC™ MGIT™ 960 system (Becton Dickinson Diagnostic Systems, Sparks, MD, USA) for INH, RIF, EMB, STP, and PZA.

For isolates recovered in the year of 2005, DST to CAP, KAN, AMK and OFX was performed by the radiometric method BACTEC 460™, using the following critical concentrations: 1.25, 5.0, 1.0 and 2.0 mg/L, respectively. For isolates after 2005, DST to CAP, AMK and OFX was performed by the BACTEC™ MGIT™ 960 system using the recommended critical concentrations for second line drugs.[483]

All strains were genotyped by 12-*loci* MIRU-VNTR as described by Supply *et al* as it was the standard genotyping technique implemented in the laboratory during the years of 2005-2007.[57]

Mutational analysis of drug-resistance associated genes

A 981 bp fragment, containing *tlyA* gene, was amplified by PCR using *tlyA*F3 (5'-AAGGCATCGCACGTCGTCTTTCC-3') and *tlyA*R3 (5´-TGTCGCCCAATACTTTTTCTACGC-3´) primers. Cycling conditions were as follows: denaturation at 94ºC for 4 min; 35 cycles of denaturation at 94ºC for 1 min, annealing at 58ºC for 1 min and extension at 72ºC for 2 min and 30 sec; lastly, a final extension step at 72ºC for 10 min was performed.

The region comprehended between the positions 1151-1590 of the *rrs* gene was amplified by PCR using primers RRS2-F (5´-TGCCGGGGTCAACTCGGAGG-3´) and RRS2-R (5´-GAACCCCTCACGGCCTACGC-3´). Cycling conditions were as follows: denaturation at 94ºC for 4 min; 35 cycles of denaturation at 94ºC for 1 min, annealing at 58ºC for 1 min and extension at 72ºC for 2 min and 30 sec; lastly, a final extension step at 72ºC for 10 min was performed.

A 601 bp fragment, containing the promoter region of *eis* gene, was amplified from all isolates by PCR using *eis*F1 (5´-GCCATGGGACCGGTACTTGC-3´) and *eis*R1 (5´-GTAGATGCCGCCCTCGCTAG-3´) oligonucleotide primers. Cycling conditions were as follows: denaturation at 94ºC for 5 minutes; 35 cycles of denaturation at 94ºC for 1 min, annealing at 56ºC for 1 min and extension at 72ºC for 1 min and 30 sec; and a final extension step at 72ºC for 10 min.

An internal 320bp (78-397) fragment of the *gyrA* gene was amplified by PCR with oligonucleotide primers *gyrA*-F (5´-CAGCTACATCGACTATGCGA-3') and *gyrA*-R (5´-GGGCTTCGGTGTACCTCAT-3'). Cycling conditions were as follows: denaturation at 94ºC for 4 min; 35 cycles of denaturation at 94ºC for 1 min, annealing at 58ºC (55ºC for *gyrA*) for 1 min and extension at 72ºC for 2 min and 30s (1 min for *gyrA*); lastly, a final extension step at 72ºC for 10 min was performed. Crude DNA extracts of each isolate were used in all molecular biology reactions. Prior to sequencing, the amplified fragments were purified using the MSB™ Spin PCRapace (Invitek™). Sequencing reactions and capillary electrophoresis were outsourced to Macrogen Inc using primers *eis*F1, *RRS*-2R, *gyrA*-F, t*lyA*F1 (5'-GTTGTTGGCCGCCCTGGAGT-3') and t*lyA*R1 (5´-GGTCTCGGTGGCTTCGTCGC-3´), as appropriate.

AMK/KAN Minimal Inhibitory Concentration

AMK and KAN minimal inhibitory concentration (MIC) for selected isolates was determined using the Resazurin Microplate Assay (REMA) system as described by Martin *et al*.[484] Selected isolates were tested in duplicate for growth at serial two-fold dilutions of both antibacillary drugs at a concentration range between 80 and 0.08 mg/L. Growth controls with no drug and a sterility control were included in each plate assay.[485]

5.4 RESULTS

Mutations in second-line drug resistance associated genes

In the present study, 44 MDR/XDR-TB strains were analyzed, 5 of which were isolated during the MDR-TB outbreak identified in the 90s. For all isolates, we have analyzed four genes associated with XDR criteria of resistance, injectable second-line drugs and FQs - *rrs*, *tlyA*, *eis*, and *gyrA*.

In *rrs* gene, six different mutations were found, including two insertions (Table 5.1). The most prevalent mutation found in *rrs* gene was A1401G (13/40), a classical mutation associated with KAN, AMK and CAP resistance, which was highly associated with the the Q1 cluster as previously reported.[25] This mutation was also found in one Lisboa3 isolate from 1998.

The mutational analysis of *tlyA* has revealed four different mutations: one missense mutation, one nonsense mutation and two frameshift mutations. All mutations detected in *tlyA* gene were detected in Lisboa3 cluster only, the most prevalent being a frameshift GT insertion at position 755 found in 13 isolates.

The screening of *eis* promoter region, more recently shown to be associated with KAN resistance, resulted in the finding of one mutation only, G-10A, previously shown by Zaunbrecher *et al* to cause overexpression of the *eis* gene.[372] This mutation was found in Lisboa3 isolates (19/27).

Table 5.1- Resistance patterns, clusters and mutations found in the 44 analyzed isolates.

Isolate	Year	First-Line Resistance Profile[a]	Second-Line Resistance[b]				Mutations[c]				MIRU-VNTR Cluster
			KAN	AMK	CAP	OFX	*eis*	*tlyA*	*rrs*	gyrA	
BK103	2005	IRSEP	R	S	R	R	G-10A	A205C (K69Q)	none	A281G (D94G)	Lisboa3
BK232	2005	IRSEP	R	R	R	R	G-10A	G223T (E75X)	none	A281G (D94G)	Lisboa3
BK643	2005	IRSEP	R	S	R	R	G-10A	Ins521CG	none	A281G (D94G)	Lisboa3
FF359	1998	IRS	nd	nd	nd	nd	G-10A	none	A1401G	A281G (D94G)	Lisboa3
HPV129	2009	IRSEP	S	S	S	R	G-10A	none	none	A281G (D94G)	Lisboa3
HPV108	2009	IRSEP	S	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
FF505	1997	IRSE	nd	nd	nd	nd	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK530	2005	IRSEP	S	S	R	R	G-10A	Ins755GT	Ins1211G	T271C (S91P)	Lisboa3
BK872	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	Ins1512G	T271C (S91P)	Lisboa3
BK30	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK81	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK127	2005	IRSEP	R	R	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK214	2005	IRSEP	R	R	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK309	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK548	2005	IRS	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK801	2005	IRSE	S	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK839	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK1144	2005	IRSEP	R	S	R	R	G-10A	Ins755GT	none	T271C (S91P)	Lisboa3
BK128	2005	IRSEP	R	R	R	R	G-10A	G223T (E75X)	G1158T	A281G (D94G)	Lisboa3
IHMT134	2009	IRSP	nd	S	S	S	none	none	none	C249G	Lisboa3
FF181	1997	IRS	nd	nd	nd	nd	none	none	none	none	Lisboa3
FF291	1998	IRS	nd	nd	nd	nd	none	none	none	none	Lisboa3
FF593	1997	IRS	nd	nd	nd	nd	none	none	none	none	Lisboa3
FF164	1997	IRS	nd	nd	nd	nd	none	none	none	none	Lisboa3
IHMT82	2009	IRS	nd	S	R	S	none	none	none	none	Lisboa3
BK519	2005	IRESP	R	S	R	R	none	none	C1402A; G1484T	A281G (D94G)	Lisboa3
BK87	2005	IRS	R	R	R	R	none	none	G1484T	A281G (D94G)	Lisboa3
HVNG1	2008	IRSEP	R	R	R	S	none	none	A1401G	A209G (H70R)	NC
FF291	1997	IRE	nd	nd	nd	nd	none	none	none	G280A(D94N)	NC
HVNG2	2009	IRS	S	S	S	S	none	none	none	none	NC
HPV113	2008	IRSEP	S	S	S	S	none	none	none	none	NC
IHMT149	2009	IRSEP	nd	R	R	R	none	none	A1401G	A281C (D94A)	Q1
IHMT151	2009	IRSEP	nd	R	R	R	none	none	A1401G	A281C (D94A)	Q1
HPV115	2008	IRSEP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
IHMT361	2008	IREP	nd	S	R	S	none	none	none	none	Q1
BK104	2005	IREP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK269	2005	IREP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK364	2005	IRSEP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK460	2005	IREP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK474	2005	IRSEP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK667	2005	IREP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK733	2005	IREP	R	R	R	R	none	none	A1401G	A281C (D94A)	Q1
BK396	2005	IRSEP	R	R	R	R	none	none	A1401G; G1158T	A281C (D94A)	Q1
BK372	2005	IR	R	R	R	R	none	none	none	none	Q1

[a]First-Line DST, I – isoniazid; R – rifampicin; E – ethambutol; S – streptomycin; P – pyrazinamide.
[b]Second-line Resistance, R – Resistant, S – Susceptible, nd – not determined.
[c]Nucleotidic mutations and, in brackets, aminoacidic mutations.

Comparing the AMK and KAN phenotypic resistance profiles we have observed that from the 17 isolates with second-line DST data bearing the *eis* G-10A mutation, nine were resistant to KAN alone; four were resistant to both KAN and AMK and four were susceptible to either (Table 5.1).

This observation opened the question concerning the molecular basis of AMK resistance when no *rrs* mutations are found. According to the published work by Zaunbrecher *et al*, *eis* promoter mutations may cause resistance to KAN but not AMK.[372] We have found however that isolates bearing *eis* G-10A mutations might be KAN and AMK resistant, KAN resistant only or susceptible to both. One hypothesis is that the level of KAN or AMK resistance is lower than the level of resistance conferred by *rrs* mutations. This low-level resistance if close enough of the tested critical concentration might yield discrepant results independently of it's genotype but dependent on the methodology employed in drug susceptibility testing. To investigate this hypothesis, we have determined the KAN and AMK MIC for selected isolates bearing different mutations using the REMA methodology (Table 5.2). We have found that isolates bearing the *rrs* A1401G mutations all exhibited a high-level resistance (>80 mg/L) to both AMK/KAN. The results obtained for the two isolates bearing the *eis* G-10A mutation yielded slightly different results, with one isolate exhibiting an MIC of 5 mg/L for KAN (equal to the critical concentration used in the MGIT™ 960 system) and 0.63 mg/L (below the critical concentration used in the MGIT™ 960 system) for AMK. The other isolate exhibited KAN and AMK MICs (10 and 1.25 mg/L, respectively) above the critical concentrations used in KAN/AMK DST (Table 5.2). Both isolates initially tested susceptible to AMK/KAN (Table 5.1), which may corroborate the proposed hypothesis that *eis* G-10A mutations may cause low-level resistance and discrepant DST results, especially in the case of AMK DST. It is important to stress the fact that the REMA methodology is susceptible to phenotypic variations and presents difficulties in obtaining accurate readings at low level concentrations.[483] Nevertheless, in all plate assays the *M. tuberculosis* H37Rv was used as a susceptible control strain and used to establish the baseline MIC for comparison of resistance levels.

In the *gyrA* gene, four different mutations were found, three of which occurring in codon 94. All mutations detected in this gene were missense mutations and were cluster associated. D94A mutation occurred only in Q1 cluster, whereas S91P and D94G were associated only with the Lisboa3 cluster. A non-clustered isolate recovered in 1997 with an unknown second-line resistance profile was found to have mutation D94N, previously associated with FQ resistance, although no second-line susceptibility results were available for this strain. The H70R mutation

found in another non-clustered isolate (HVNG1) could not be associated with FQ resistance since the isolate is susceptible to OFX.

Table 5.2 – KAN and AMK MIC for selected strains bearing the rrs A1401G and eis G-10A mutations.

Isolate	Year	Mutations		MIC (mg/L)	
		rrs	*eis*	KAN	AMK
H37Rv	-	none	none	1.25	0.63
HPV129	2009	none	G-10A	10	1.25
HPV108	2009	none	G-10A	5	0.63
HPV115	2008	A1401G	none	>80	>80
HVNG1	2008	A1401G	none	>80	>80
IHMT149	2009	A1401G	none	>80	>80
IHMT151	2009	A1401G	none	>80	>80

Genotypic comparison reveals clonality and resistance acquisition dynamics.

The overall isolate distribution through Q1 and Lisboa3 clusters allowed the following observations: in Q1 cluster (except for one isolate) one mutation only (*rrs* A1401G) ensures resistance to all three second-line injectable drugs, with no mutation detected in *eis* and *tlyA* genes. Moreover, in Q1 cluster two mutational events, in *rrs* and *gyrA*, were enough for a MDR-TB strain to become XDR in two mutational steps. With the current panel of isolates it is impossible to determine which resistance was acquired in the first place since all isolates with the *rrs* A1401G mutation also have the *gyrA* D94A mutation.

In Lisboa3 cluster, however, it is possible to establish the dynamics of resistance acquisition. The mutational analysis itself allows the discrimination of the entire cluster as shown by the intra-cluster mutational diversity. We have observed that whenever a mutation was present in *tlyA* gene, no classical *rrs* mutation (A1401G, C1402A/T or G1484T) was detected and, that isolates belonging to Lisboa3 cluster with a *tlyA* mutation also had an *eis* mutation (G-10A) (Table 5.1).

5.5 DISCUSSION

In the present study, we have analyzed the second-line drug resistance associated genes in 44 MDR/XDR-TB clinical isolates, of which 40 belong to two different highly-prevalence clusters. Lisboa3 cluster was first described in the 90s already associated with MDR, and later XDR, while Q1 cluster was described in 2008 as a more homogeneous cluster, from a mutational standpoint and, already associated with MDR and XDR.[24-26]

Looking at the second-line resistance profiles, specifically, resistance to KAN, AMK and CAP, we found that in Q1 cluster (except for one isolate) the acquisition of resistance to these three injectable second-line drugs was caused by a single mutational event (*rrs* A1401G); however, in Lisboa3 cluster our data suggests that resistance to KAN/AMK was first acquired through an *eis* hypermorphic mutation and, at a later stage to CAP due to *tlyA* mutations. Therefore, both genetic clusters illustrate single and multistep processes of resistance development by strains that were subsequently transmitted in the community generating primary XDR-TB cases. Such observations allow the establishment of a putative model for the resistance acquisition pattern in Lisboa3 cluster (Figure 5.1). This model consists in a multistep process more extensive than the one observed in the Q1 cluster. We propose that in Lisboa3 cluster an MDR strain initially acquired resistance to KAN, followed by FQ resistance acquisition, twice and independently, followed by CAP resistance acquisition through multiple events and finally, acquisition of a higher level of resistance to AMK, KAN and CAP through an *rrs* A1401G mutation. The only strain (FF359) to exhibit this later step of resistance acquisition was isolated in 1998 and its dissemination was most probably contained since no Lisboa3 isolate with a *rrs* A1401G was found in recent years. For the Lisboa3 strains bearing the *gyrA* S91P and *tlyA* Ins755GT mutations it is impossible to determine if FQ resistance was acquired first than CAP resistance or vice-versa since all strains with one of these mutations also have the other. Noteworthy is the fact that 2 out of 7 strains isolated between 1997-1998 and involved in the original Lisbon MDR-TB outbreak described by us, already had at the time the mutational profile of recent circulating strains.[26] Second-line DST is not available for these strains but the fact that some isolates exhibit the Lisboa3 characteristic *rrs*, *tlyA* and *gyrA* mutations indicate that they were most likely XDR-TB strains at that time. This highlights the ancient XDR-TB problem that is present in Lisbon Health Region for almost two decades.

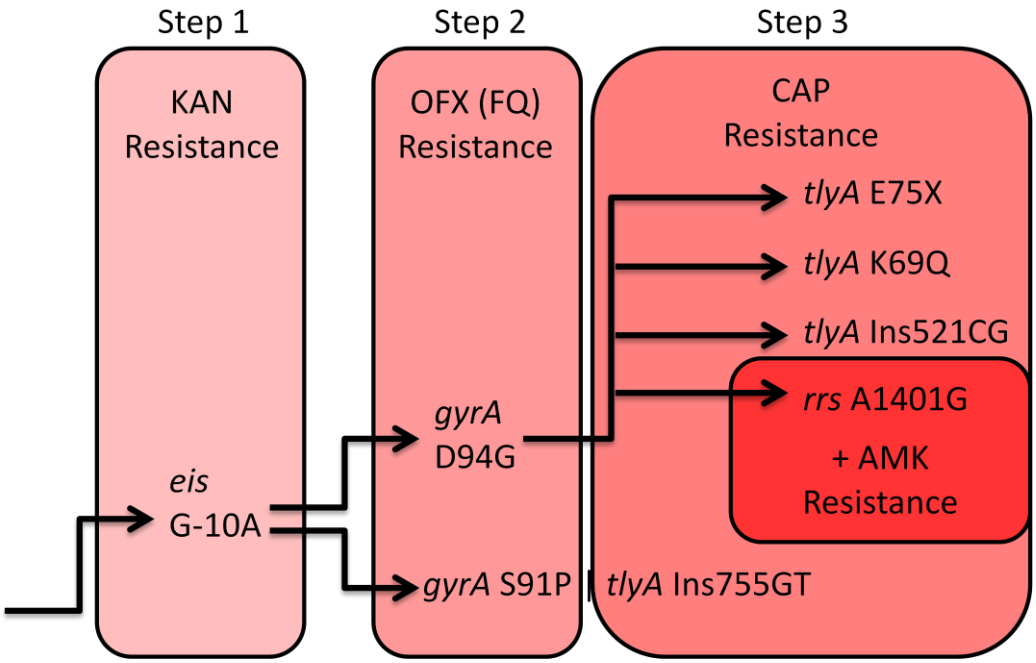


Figure 5.1 – Multistep process of resistance acquisition dynamics in Lisboa3 cluster. The scheme represents the process through which XDR has most likely been acquired multiple independent times in a maximum of three steps. Step 1 consisted in the acquisition of KAN low-level resistance through an *eis* G-10A mutation, followed by the acquisition of FQ resistance by a *gyrA* D94G or S91P mutation, although regarding the strains with the S91P mutation it is impossible to ascertain whether FQ resistance or CAP resistance was acquired in the first place. Strains with the *gyrA* D94G mutation acquired CAP resistance multiple times through different mutations in tlyA gene. One strain was found to have acquired CAP and KAN resistance through a *rrs* A1401G mutation, thus strengthening the KAN resistance.

The pattern and the proposed resistance acquisition dynamics described above illustrate the need for a more efficient MDR/XDR-TB management in two different aspects. First, the need to contain specific clusters or families of strains associated with MDR/XDR-TB, such containment should be supported in a better contact tracing for affected individuals and more rapid and reliable diagnostics. Given the time-consumable culture and DST of *M. tuberculosis*, several rapid molecular diagnosis systems have become available. However, MDR-TB molecular diagnosis *per se* is of no use if no information regarding susceptibility to second-line drugs is available. Starting an empirical or standardized treatment for MDR-TB may result in further drug resistance development, and such is highlighted by the dynamics shown in the present study, in which two mutational events are enough for a strain to become XDR. The necessary tools to perform such molecular testing in a setting-adapted approach are sometimes not readily available commercially and still, before implementation of these commercial tests it is necessary to determine the local applicability due to the geographical

variability of resistance-associated mutations. As the body of knowledge concerning new mutations and drug resistance associated genes expand it will be gradually easier to associate both genotype and phenotype.

Another aspect to consider is the issue of mismanagement of drug therapy, poor regimen implementation, poor compliance or default that leads to cumulative resistance acquisition. Although the high prevalence of XDR-TB in the region illustrates the need for new drugs and new drug regimens for MDR/XDR-TB treatment, special care must be taken in the introduction of new drugs in TB treatment.[486] The additional emergence of resistance to new compounds may render these drugs ineffective in a short time period.[487, 488]

Zaunbrecher *et al* proposed that *eis* promoter mutations would be associated with KAN resistance only, and therefore explain the lack of cross resistance between KAN and AMK in some clinical isolates.[372] Although in such cases an increase in the level of AMK resistance was observed, the strains were phenotipically considered susceptible. The question resides in the degree of AMK resistance conferred by *eis* promoter mutations has clinical relevance and impairs chemotherapy and can be detected in drug susceptibility testing? Our data shows that four isolates containing *eis* G-10A mutation and no *rrs* mutations were phenotipically resistant to AMK and KAN. Is it therefore possible that the degree of AMK resistance may be close enough to the critical concentration used in drug susceptibility testing resulting in undetectable low-level resistance?

Furthermore, we have determined the KAN and AMK MIC for selected strains and the data obtained corroborate the hypothesis that the increase in the KAN MIC for *eis* G-10A isolates is accompanied by a slighter AMK MIC increase. A two-fold increase in the AMK MIC, when compared with the AMK MIC for *M. tuberculosis* H37Rv, is not as significant as the increase in KAN MIC but, it may be sufficient to confer border-line/undetectable resistance and hence, explain discrepant or inconsistent results. The same may happen with KAN DST, as one strain exhibited a four-fold increase and another strain an eight-fold increase in KAN MIC. Supporting this, Engstrom *et al* have reported the occurrence of *eis* G-10A mutations among KAN susceptible isolates, which may be due to the low-level resistance conferring MICs close to the critical concentration used in DST.[366] Similarly, Campbell *et al* reported that in 111 KAN-resistant isolates, 32% had *eis* mutations with no isolate accumulating mutations at *rrs* and *eis* simultaneously.[219] Moreover, the authors also reported one isolate resistant to KAN and AMK with an *eis* promoter mutation only. In one other study by Feuerriegel *et al*, 10 AMK and CAP resistant isolates did not show any mutation in the *rrs* and *tlyA* genes and the same was true

for strains resistant to AMK alone.[367] One possible explanation is that AMK resistance in these isolates is mediated by *eis* promoter mutations. The authors however have not tested the isolates for susceptibility to KAN, which would help to clarify this mechanism. In fact, current directives from Portuguese Health Authorities state that second-line DST is not mandatory for KAN, only for AMK and CAP. Nonetheless, as confirmed by this study, it is important to stress that susceptibility to AMK does not imply susceptibility to KAN. In our opinion, KAN DST is of great importance as KAN resistance alone may be associated with undetectable AMK low-level resistance, which may eventually impact on the clinical outcome or resistance development. The non-simultaneous occurrence of mutations at *eis* and *rrs* in more recent isolates, representing a genetic cluster (Lisboa3) circulating for over a decade, suggests that mutations in either of these genes are sufficient to confer phenotypic resistance in the host and impair clinical treatment. Otherwise, we would expect to find at this stage Lisboa3 strains that have acquired AMK and KAN high-level resistance.

Also recently, data from Kim *et al* show that the Eis protein has an important role in the suppression of the host immune system through the acetylation of DUSP16/MKP-7, inhibiting inhibition of JNK-dependent autophagy, phagosome maturation, and reactive oxygen species generation.[489] This Eis important immune modulator property combined with its overexpression might be responsible for an increased virulence phenotype and be a clue to explain the high prevalence and persistence of certain Lisboa family strains in the region.

5.6 ACKNOWLEDGEMENTS

The authors wish to thank to the laboratory personnel of Hospital Pulido Valente, Instituto de Higiene e Medicina Tropical, Hospital de S. Bernardo, Hospital de St. Marta, Hospital S. João de Deus, Hospital Curry Cabral, Hospital Fernando da Fonseca, Hospital Sta. Maria, Hospital de Torres Vedras, and Laboratório de Pneumologia da Sub-região de Saúde de Lisboa, for data collection and processing; and to the staff of Unidade de Micobactérias of Centro de Bacteriologia, Instituto Nacional de Saúde Dr. Ricardo Jorge, whose high quality of service and cooperation have made this work possible.

5.7 FUNDING

This work was partially supported by Project Ref. SDH49: “Early Molecular Detection of M/XDRTB in the Great Lisbon Healthcare Region” from Fundação Calouste Gulbenkian (FCG, Portugal). J. Perdigão, D. Machado and C. Silva were supported by FCT grants SFRH/BD/45388/2008, SFRH/BD/65060/2009 and SFRH/BD/73579/2010, respectively.

5.8 TRANSPARENCY DECLARATIONS

None to declare

CHAPTER 6

GidB mutation as a phylogenetic marker for Q1 cluster *Mycobacterium tuberculosis* isolates and intermediate-level streptomycin resistance determinant in Lisbon, Portugal

João Perdigão[1], Rita Macedo[1,2], Diana Machado[3], Carla Silva[1], Luísa Jordão[4], Isabel Couto[3,5], Miguel Viveiros[3], and Isabel Portugal[1]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Public Health Laboratory: Mycobacteriology/Tuberculosis, Public Health Department, Administração Regional de Saúde de Lisboa e Vale do Tejo, I.P., Lisboa, Portugal; [3]Grupo de Micobactérias, Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (IHMT/UNL), Lisboa, Portugal; [4]Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisboa, Portugal; [5]Centro de Recursos Microbiológicos (CREM), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal.

Manuscript in preparation

6.1 ABSTRACT

Development of STP-resistance in *M. tuberculosis* is usually associated with mutations in *rpsL* and *rrs* genes, although up to 50% of clinical STP-resistant isolates may present no mutation in either of these genes. In the present report we investigate the role of *gidB* gene mutations in STP resistance.

We have analyzed 52 STP-resistant and 30 STP-susceptible *M. tuberculosis* clinical isolates by sequencing and endonuclease analysis of the *gidB* and *rpsL* genes. All clinical isolates were genotyped by 12-*loci* MIRU-VNTR.

The *gidB* gene of 18 STP-resistant isolates was sequenced and four missense mutations were found: F12L (1/18), L16R (18/18), A80P (4/18) and S100F (18/18). The remaining isolates were screened by endonuclease analysis for mutations A80P in *gidB* and K43R in *rpsL* gene. Overall, mutation A80P in *gidB* gene was found in 7 STP-resistant isolates and 12 STP-susceptible multidrug resistant isolates. Also noteworthy, is the fact that *gidB* mutations were only present in isolates without *rpsL* and *rrs* mutations, all from genetic cluster Q1. STP semi-quantitative drug susceptibility testing showed that isolates carrying the *gidB* A80P mutation were STP intermediate-level resistant and that standard drug susceptibility testing yielded inconsistent results probably due to borderline resistance.

We conclude that *gidB* mutations may explain the high number of STP-resistant strains with no mutation in *rpsL* or *rrs*. These mutations might occasionally confer undetected STP low-level resistance in regular drug susceptibility testing. Also, GidB A80P mutations may serve as surrogate markers for Q1 cluster isolates that are associated with MDR/XDR-TB.

6.2 INTRODUCTION

STP is an aminoglycoside antibiotic, the first that was introduced in the treatment of TB, revolutionizing the treatment of this disease. Although mostly regarded as a tuberculostatic agent, STP is no longer a first option in TB standardized treatments for new cases. It is still incorporated as first-line TB therapeutical regimens in patients that have previously been treated for TB (former WHO Category II ''retreatment'' regimen) and in some countries incorporated in second line regimens for drug-resistant TB, although it's not the first choice for injectable aminoglycosides given the high rates of STP resistance in drug-resistant TB.[203, 490]

In spite of this, due to the global emergency of resistant *M. tuberculosis* strains, such as MDR and XDR strains, STP may regain an important role in TB management.

Portugal has an intermediate incidence rate when compared with other countries of the European Union with 2 398 new cases (22.5 new cases per 100 000 habitants), in 2010, but a worrying situation regarding M/XDR-TB.[21, 24-26, 482] The latest TB surveillance data available and published by the ECDC for STP resistance reported, in 2008, a 9.5% prevalence of STP resistance (156 cases out of 1641 cases with DST results), which exceeded other antibacillary drug resistance prevalence's such as INH (7.4%) and RIF (1.8%). [491]

STP inhibits protein synthesis through irreversible binding to the A-site of the 30S ribosomal subunit and to the ribosomal protein S12. Ribosomes combined with STP are unable to initiate and elongate during the process of gene translation. STP also prevents dissociation of ribosomal 50S and 30S subunits. [312, 323]

Accordingly with its mode of action, STP resistance is usually mediated by mutations in *rpsL* and *rrs* genes, which encode the ribosomal protein S12 and the 16S rRNA, respectively. [312] The most common mutation occurs in codon 43 of the *rpsL* gene and can be detected by PCR-RFLP with *Mbo*II.[463] Mutations conferring STP resistance in *rrs* have also been described, usually in the 530 and 912 loop regions, and since *M. tuberculosis* only possesses one copy of the rRNA operon, one mutation in this gene is enough to cause resistance.[312, 492] More recently, a new gene, *gidB*, encoding a rRNA methyltransferase was implicated in STP resistance.[325] The gene product of *gidB* is involved in the methylation of 16S rRNA, more specifically, at position G527. Since binding of STP to 16S rRNA is crucial to its action mechanism, mutations in *gidB* may well be responsible and explain STP resistance when no mutation is found on *rpsL* or *rrs*.[325] Notably, in the STP-resistant strains isolated in our laboratory, *rrs* mutations that confer STP resistance were never detected in the STP-resistant strains lacking *rpsL* mutations.[24]

The *gidB* gene is highly conserved in all eubacteria, including *Mycoplasma genitalium,* and loss of function has been implicated in high-level STP resistance in *Salmonella* but with low-level resistance in *M. tuberculosis*.[435, 493]

In this study, we screened 52 STP-resistant and 30 STP-susceptible isolates circulating in Lisbon Health Region for mutations in *gidB* gene and its association with STP resistance.

6.3 METHODS

Clinical Isolates

Two sets of isolates were analyzed in the present study, the first comprised by a total of 82 *M. tuberculosis* isolates, 52 STP-resistant isolates and 30 STP-susceptible (14 multidrug-resistant and 16 susceptible to all first-line drugs), recovered from the same number of patients between 2004-2006. For this set of isolates, only DNA was available. All isolates were collected from hospitals and public health laboratories in Lisbon Health Region.

The second set of isolates is comprised by 6 isolates, 3 STP-resistant and 3 STP-susceptible, and was included in the study to provide data on STP semi-quantitative Drug Susceptibility Testing (qDST).

STP susceptibility testing

All isolates were tested for STP susceptibility by the fluorimetric BACTEC™ MGIT™ 960 system (Becton Dickinson Diagnostic Systems, Sparks, MD, USA) using a critical STP concentration of 1,0 mg/L, according to the manufacturer's instructions.

STP qDST was carried out for the selected isolates using three STP concentrations (1.0, 4.0 and 20.0 mg/L) on the MGIT 960 system and the Epicenter V5.80A software equipped with the TB eXIST module (Becton Dickinson Diagnostic Systems™, Sparks, MD, USA) as previously described.[151]

PCR amplification and DNA sequencing

The ORF of *gidB*, plus 96 nucleotides upstream and 56 nucleotides downstream, was amplified by PCR resulting in a 828 bp fragment using forward oligonucleotide primer gidBF1, and reverse oligonucleotide primer gidbR1 (Table 6.1). Cycling conditions for this amplification consisted of an initial denaturation step at 94ºC for 10 min; 40 cycles of a denaturation step at

94ºC for 1 min, a primer annealing step at 67.5ºC for 35 sec and, an extension step at 72ºC for 1 min; a final extension step at 72ºC for 10 min was performed.

The entire *rpsL* ORF, plus 46 nucleotides upstream and 86 nucleotides downstream, was also amplified by PCR resulting in a 504 bp fragment using forward oligonucleotide primer RPSL-1, and reverse oligonucleotide primer RPSL-2 (Table 6.1).[450] Cycling conditions for this amplification consisted of 40 cycles of a denaturation step at 94ºC for 1 min, a primer annealing step at 57ºC for 2 min and, an extension step at 72ºC for 2 min.

The region between nucleotides 10-1037 of the *rrs* gene was amplified by PCR using forward oligonucleotide primer RRS-1 and reverse oligonucleotide primer RRS-2 (Table 6.1).[450] Cycling conditions consisted in 40 cycles of a denaturation step at 94ºC for 1 min, primer annealing step at 60ºC for 2 min, and an extension step at 72ºC for 2 min.

Amplicons were purified with Wizard™ SV Gel and PCR Clean-Up System (Promega™) prior to sequencing. Sequencing reactions were performed with BigDye Terminator Cycle Sequencing Kit with AmpliTaq DNA polymerase (Applied Biosystems™) using oligonucleotide primers: RPSL-1 for *rpsL* amplicons, RRS-1 and RRS-2 for *rrs* amplicons and, gidBF1 and gidBR1 for *gidB* amplicons (Table 6.1).

Table 6.1 - Oligonucleotides used for amplification and sequencing of the *rpsL* and *gidB* genes.

	Sequence	**Tm (ºC)**	**Reference**
RPSL-1	5'-GGCCGACAAACAGAACGT-3'	63.8	Kirschner *et al* [450]
RPSL-2	5'-GTTCACCAACTGGGTGAC-3'	58.8	Kirschner *et al* [450]
RRS-1	5'-GAGAGTTTGATCCTGGCT-3'	62.9	Kirschner *et al* [450]
RRS-2	5'-TGCACACAGGCCACAAGG-3'	72.5	Kirschner *et al* [450]
gidBF1	5´-AAGCGATGCGTGGCCGAGCG-3´	69.4	This study
gidBR1	5´-CTGGCCCGACCTTACGAGCG-3´	68.4	This study

PCR-RFLP analysis

Mutation A128G in *rpsL* was detected by *Mbo*II (New England Biolabs™) digestion of the *rpsL* amplicon. Two micrograms of PCR product were digested with 5U of *Mbo*II accordingly to the manufacturer's instructions for 1 h at 37ºC. Absence of restriction indicates the presence of mutation A128G.

Mutation G238C in *gidB* was detected as above with the use of *Bsr*DI enzyme. The occurrence of restriction indicates the presence of mutation G238C.

MIRU-VNTR genotyping

All isolates were genotyped by 12-*loci* MIRU-VNTR as described by Supply *et al*.[57] A dendrogram was constructed at MIRU-VNTRplus web application available at http://www.miru-vntrplus.org using the Dsw measure of genetic distance and the UPGMA. A cluster was defined as group of more than one isolate sharing the same MIRU-VNTR profile.

Bioinformatic analysis of GidB mutations

Alignment of GidB aminoacidic sequences was carried out using BioEdit (v.7.0.5.2, T.A. Hall) for orthologous sequences from *M. tuberculosis* H37Rv (GI: 15611055), F11 (GI: 148825127), CDC1551 (GI: 15843553); *Mycobacterium bovis* AF2122/97 (GI: 31795092); *Mycobacterium africanum* GM041182 (GI: 339633908); *Mycobacterium smegmatis* MC^2155 (GI: 118471658); *Mycobacterium leprae* TN (GI: 15828466); *Mycobacterium avium* subsp. *avium* ATCC 25291 (GI: 254777657); *Mycobacterium kansasii* ATCC 12478 (GI: 240168403); *Corynebacterium diphtheriae* NCTC 13129 (GI: 38234911); *Arthrobacter globiformis* NBRC 12137 (GI: 359776049); *Actinomyces odontolyticus* F0309 (GI: 293191006); *Nocardia brasiliensis* ATCC 700358 (GI: 378817268); *Proteus mirabilis* HI4320 (GI: 197286881); *Yersinia pestis* KIM10+ (GI: 22128011); *Escherichia coli* MS 21-1 (GI: 300940926) and O157:H7 str. EDL933 (GI: 15804340); *Staphylococcus pettenkoferi* VCU012 (GI: 365225940); *Bacillus subtilis* subsp. *subtilis* str. 168 (GI: 16081152); *Mycoplasma genitalium* G37 (GI: 12045241); *Haemophilus influenzae* Rd KW20 (GI: 16272434); *Streptococcus pneumoniae* D39 (GI: 116516833); *Enterococcus faecium* PC4.1 (GI: 293379357); *Neisseria meningitidis* alpha522 (GI: 389604808); *Brucella melitensis* ATCC 23457 (GI: 225853491) and, *Salmonella enterica* subsp. *enterica* serovar Typhi str. M223 (GI: 213850448).

Impact of specific mutations on protein function was assessed based on the degree of conservation of amino acid residues using the SIFT program server available at

http://sift.jcvi.org/. Comparison was made using the GidB sequence from *M. tuberculosis* H37Rv against the UniRef90 database.[494]

6.4 RESULTS

Screening a set of 18 STP-resistant isolates for *gidB* mutations we have found four different missense mutations: F12L, L16R, A80P and S100F. Mutations L16R and S100F were present in all isolates tested. When analyzing the *gidB* homologous genes from both *M. tuberculosis* CDC1551 and F11 we found mutation S100F to be present on both strains and that *M. tuberculosis* F11 also had the L16R mutation. These two mutations were therefore considered natural polymorphisms and were not considered as mutations associated with STP resistance. GidB A80P and F12L mutations were detected in 4 and 1 isolates, respectively.

The A80P mutation was therefore considered the most prevalent mutation occurring in the *gidB* gene and a method to rapidly screen for G238C (A80P) mutation in *gidB* through PCR-RFLP was devised through amplicon digestion with *Bsr*DI. Thirty-four additional STP-resistant isolates and 30 STP-susceptible isolates were characterized by this manner. The results obtained for each isolate were compared with the characterization of the *rpsL* gene.

Overall, the GidB A80P mutation was detected in 19 out of 82 isolates, 7 STP-resistant and 11 STP susceptible. Only MDR isolates were found to bear the A80P mutation, whereas the 16 pan-susceptible isolates included in the study didn't harboured this mutation (Figure 6.1). Also noteworthy, all strains containing the GidB A80P mutation lacked *rpsL* and *rrs* mutations.

To examine the distribution of this prevalent mutation and in an attempt to associate this specific mutation with any genetic cluster, family or clade, all isolates were genotyped by 12-*loci* MIRU-VNTR. We found that all isolates carrying the GidB A80P mutation belonged to genetic cluster Q1 (Figure 6.1). The occurrence of GidB A80P mutation in all Q1 isolates and the fact that this mutation occurs among both STP-resistant and susceptible isolates may suggest that this mutation has no contribution to STP-resistance but may rather be, from a resistance stand-point, a neutral phylogenetic polymorphism characteristic of Q1 isolates. Alternatively, this mutation might constitute an adaptation mechanism acquired by an ancestral Q1 isolate that yielded clinically significant but borderline STP resistance on standard DST.

To further elucidate the contribution of GidB A80P mutation to STP resistance, we decided to study the STP qDST profiles in a second set of 6 *M. tuberculosis* clinical isolates composed by 4 Q1 isolates carrying the GidB A80P mutation, one Beijing strain carrying the *rpsL* K43R mutation and a pan-susceptible isolate with no mutation in *gidB*, *rpsL* or *rrs*. The results obtained show that the four Q1 isolates were all resistant to 1 mg/L, of which three were resistant to 4 mg/L, while the remaining isolate displayed an intermediate resistance to 4 mg/L (Table 6.2). All Q1 isolates were susceptible to 20 mg/L. The Beijing strain with the K43R mutation on the *rpsL* gene was resistant to the three tested STP concentrations whereas the pan-susceptible clinical strain and the control strain H37Rv were found to be susceptible to the three tested concentrations (Table 6.2). The results obtained show that the Q1 isolates carrying the GidB A80P mutation display an increased STP resistance-level towards an intermediate-level resistance (Table 6.2). Three Q1 isolates with the GidB A80P mutation (two STP susceptible and one resistant) were also blindly retested in duplicate, using standard DST methodology on BACTEC 960 MGIT with a STP breakpoint of 1 mg/L. Surprisingly, the two isolates (IHMT308 and 361) initially considered as STP-susceptible were considered resistant upon retesting (Table 6.2).

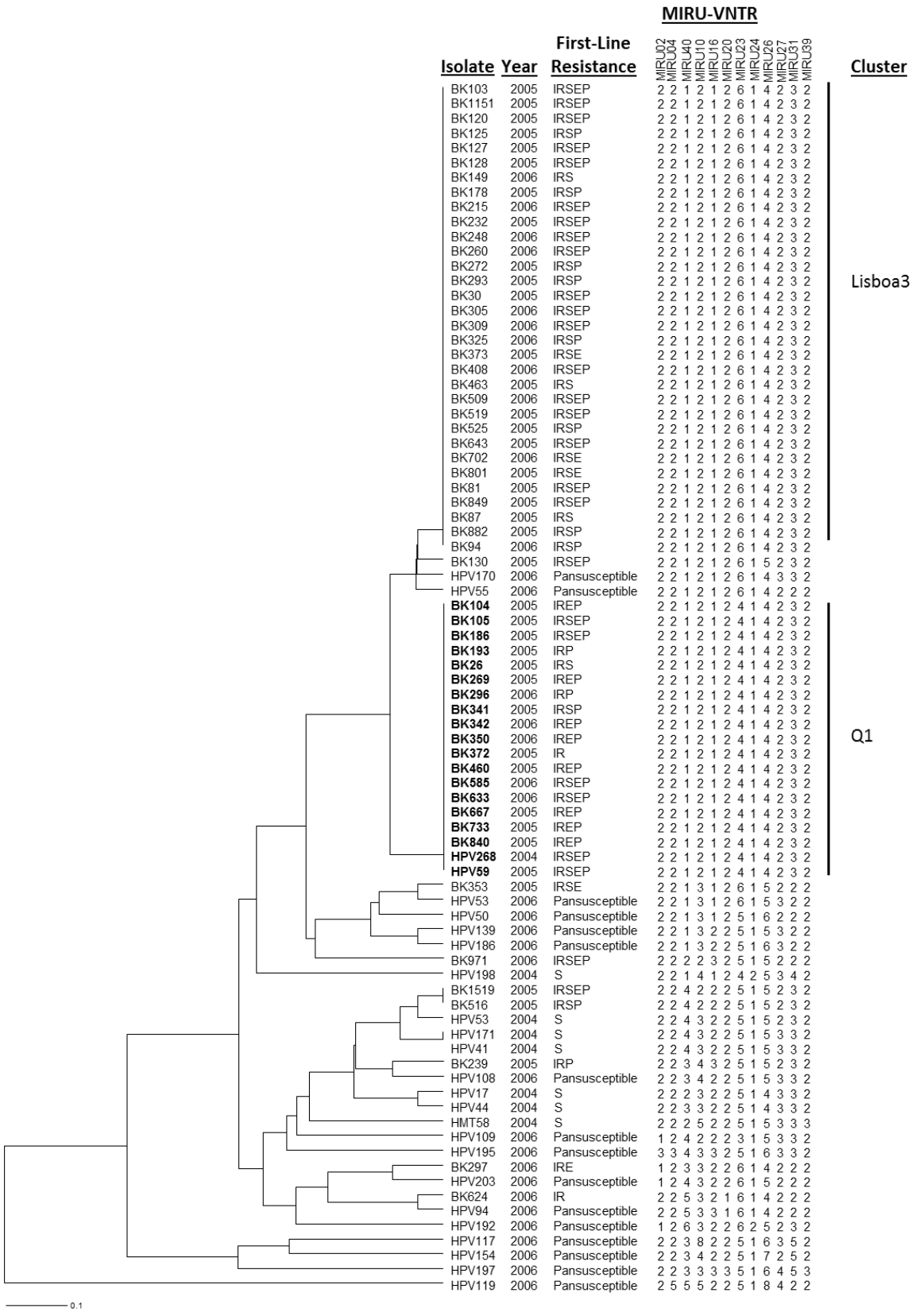


Figure 6.1– MIRU-VNTR dendrogram of the 82 *M. tuberculosis* clinical isolates screened for *gidB* and *rpsL*. Isolates carrying the GidB A80P mutation are shown in bold. Resistance profile: I, isoniazid; R, rifampicin; S, streptomycin; E, ethambutol; and P, pyrazinamide.

Table 6.2– STP qDST for selected strains carrying the GidB A80P mutation.

Isolate	Year	First-line Resistance[a]	MIRU-VNTR Cluster[b]	Mutations[c]			STP qDST[d]			STP DST Retest[e]	
				gidB	*rpsL*	*rrs*	1 mg/L	4 mg/L	20 mg/L	#1	#2
H37Rv	-	Pan-susceptible	N.C.	None	None	None	S	S	S	nd	nd
HPV65	2008	Pan-susceptible	N.C.	None	None	None	S	S	S	nd	nd
IHMT295	2008	IRSEP	N.C. (Beijing)	None	K43R	None	R	R	R	nd	nd
HPV115	2008	IRSEP	Q1	G238C (A80P)	none	None	R	R	S	nd	nd
IHMT308	2008	IRP	Q1	G238C (A80P)	None	None	R	R	S	R	R
IHMT361	2008	IREP	Q1	G238C (A80P)	None	None	R	I	S	R	R
IHMT149	2009	IRSEP	Q1	G238C (A80P)	None	None	R	R	S	R	R

[a] First-Line resistance: I – Isoniazid; R – Rifampicin; S – Streptomycin; E – Ethambutol; P - Pyrazinamide.
[b] N.C. – Non-clustered.
[c] For these isolates the respective genes have been sequenced as described in Material and Methods section. S100F and L16R mutations in *gidB* are omitted.
[d] STP qDST: S – susceptible: I – intermediate; R – Resistant.
[e] STP DST Retest: R – Resistant; S – Susceptible; nd – not determined.

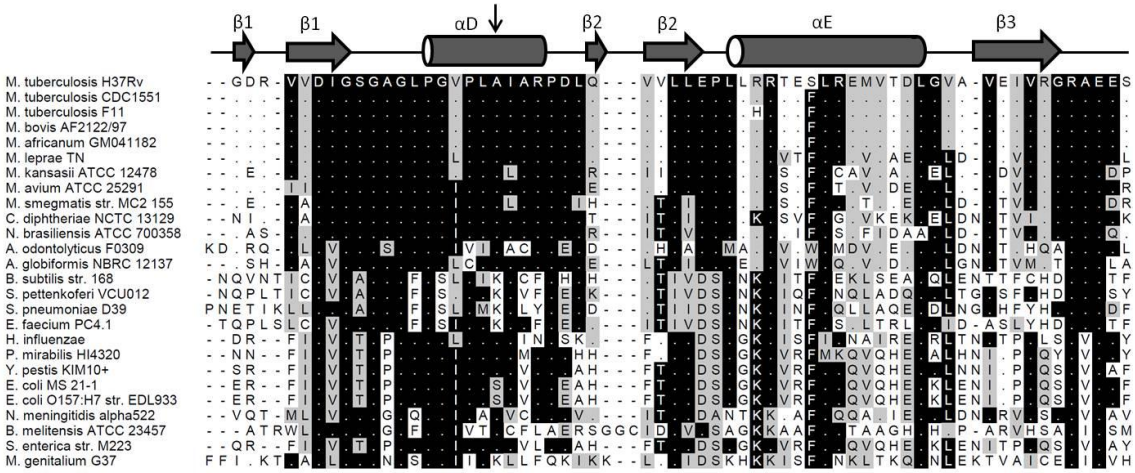


Figure 6.2 - Alignment of GidB homologous sequences from different bacterial species showing conservation of Alanine 80 (vertical arrow) and positioning in the alpha-helix D deduced from the crystal structure of *E. coli* GidB (RCSB Protein Data Bank ref. 1JSX).[495] Dots (.) represent identical amino acid residues and alignment gaps are represented by dashes (-). White, gray and black shading represent increasing conservation at each position, respectively.

Analysis of the amino acidic sequence of GidB and comparison with other GidB homologous sequences from other *M. tuberculosis* strains, other mycobacteria species and different bacterial species showed that the GidB alanine 80 is highly conserved among mycobacteria and that some degree of conservation exists among both Gram-positive and –negative bacteria, in particular among Gram-negative (Figure 6.2). Furthermore, the GidB A80P mutation was predicted to affect protein function by SIFT score analysis, yielding a SIFT score of 0.01, resulting from the analysis of 29 sequences.

6.5 DISCUSSION

In the present study, we have analyzed the *gidB* gene, encoding a rRNA methyltransferase, in STP-resistant and susceptible isolates. Methyltransferases have demonstrated a growing importance in the action mechanism and resistance pathways to several antimicrobial agents in *M. tuberculosis*. Intrinsic resistance to macrolides is thought to be associated with 23S rRNA methyltransferases encoded by the *erm* genes whereas loss of TlyA methyltransferase confers resistance to cyclic peptides, *e.g.*, CAP. However, most rRNA methyltransferases are still unknown in *M. tuberculosis.*[496]

The screening of *gidB* mutations revealed that besides two naturally occurring polymorphisms (L16R and S100F), two different missense mutations were found, F12L and A80P, the latter being the most prevalent. Both mutations had never been described before and, were only detected in isolates without *rpsL* and *rrs* mutations. As a result, *gidB* mutations may account for STP resistance in isolates with both wild-type *rpsL* and *rrs.* Nevertheless, the A80P mutation was detected in MDR isolates susceptible to STP but not in pan-susceptible isolates.

The A80P mutation has not, to our extent of knowledge, been previously described and bioinformatic analysis revealed that this mutation occurs in the alpha-helix D close to the S-adenosyl-L-methionine binding consensus sequence, according to homology analysis with *E. coli* GidB and respective crystal structure.[495] Moreover, this substitution does not appear to be chemically conservative and may eventually hamper the functionality of GidB in methylating the 16S rRNA leading to an increase in STP resistance level. This latter hypothesis was further supported by SIFT score analysis, which predicted that the A80P mutation has a probable impact on GidB function and consequently on the development of STP resistance.[325]

Other studies have reported the occurrence of *gidB* mutations in susceptible isolates and this phenomenon appears to be related with STP low-level resistance.[214, 497] STP qDST revealed that the majority of tested strains carrying the A80P mutation exhibit an intermediate STP resistance level (*i.e.*, either resistant or intermediate resistance to 4 mg/L but susceptible to 20 mg/L) which may correlate with the STP low-level resistance reported by others relatively to other *gidB* mutations in *M. tuberculosis* clinical isolates.[327, 435] It is possible that strains carrying this mutation may occasionally yield inconsistent DST results as demonstrated by the STP retesting of selected isolates. In the present study the use of the novel qDST methodology takes into account the growth rate in both drug-containing and –free media to establish intermediate resistance levels relative to a specific concentration, rather than just resistant or susceptible as in the traditional MGIT methodology.[151]

Our results indicate that a mutation in *gidB* may be enough to confer clinically relevant STP resistance, although we do not discard a possible contribution of efflux mechanisms to the overall STP resistance level as proposed by Spies *et al* when studying the synergistic effect between *gidB* mutations and efflux pump inhibitors.[326]

The association with the GidB A80P mutation with the genetic cluster Q1, a highly prevalent cluster of M/XDR-TB isolates in Lisbon, Portugal, is important since it allows the use of this mutation as a surrogate marker for Q1 isolates and can be useful for its detection. The detection of these isolates may be useful in managing M/XDR-TB in the region since early detection of Q1 isolates would prove valuable for the design of adequate treatment regimens and patient isolation.[24, 25, 482]

Other mutations may be useful for assignment to *M. tuberculosis* lineages. For example, the L16R substitution is associated with isolates from the LAM lineage, which includes the *M. tuberculosis* F11 strain.[214, 327] In fact, approximately 50% of the strains circulating in the Lisbon Health region belong to the LAM lineage of which 1.67% represent Q1 isolates belonging to the LAM4 sub-lineage (unpublished data).[498] The phylogenetic association of the GidB S100F polymorphism is, however, unclear as it was present in all the 18 isolates.

Given the high resistance rates in this and other settings, the correct assessment of STP resistance level and eventual association with a molecular marker can be useful in the decision of whether STP should be used in some therapeutic regimens.

Nevertheless, it is also important to stress that the addition of STP alone to a treatment regimen such as the former WHO recommended Category II regimen is counterproductive as it

may lead to the acquisition of STP resistance and, amplification of resistance to other antibacillary drugs in use.[203, 499] A recently published study performed in Peru by Ponce *et al* showed that the cure rate was lower for WHO Category II regimen compared to treatment with Category I regimen. According to the authors, this should support the phasing-out of WHO Category II regimen.[500] Additional data from other countries seem to also support this notion.[501, 502]

In conclusion, our results suggest that the cluster-specific *gidB* polymorphism A80P is responsible for an intermediate level of STP resistance, which has implications for routine STP DST testing as currently performed using a single critical concentration; STP prescription; and, M/XDR-TB detection. Our data, further contributes to global awareness of the role of *gidB* mutations in the process of STP resistance development.

6.6 FUNDING

This work was partially supported by Project Ref. SDH49: "Early Molecular Detection of M/XDRTB in the Great Lisbon Healthcare Region" from Fundação Calouste Gulbenkian (FCG, Portugal). J. Perdigão, D. Machado and C. Silva were supported by FCT grants SFRH/BD/45388/2008, SFRH/BD/65060/2009 and SFRH/BD/73579/2010, respectively.

6.7 COMPETING INTERESTS

The authors have nothing to declare.

6.8 ETHICAL APPROVAL

Not required.

CHAPTER 7

Unraveling *Mycobacterium tuberculosis* genomic diversity and evolution in Lisbon, Portugal, a highly drug resistant setting

João Perdigão[1], Hugo Silva[1], Diana Machado[2], Rita Macedo[3], Fernando Maltez[4], Carla Silva[1], Luisa Jordao[5], Isabel Couto[2,6], Kim Mallard[7], Francesc Coll[7], Grant A. Hill-Cawthorne[8,9], Ruth McNerney[7], Arnab Pain[8], Taane G. Clark[7], Miguel Viveiros[2] and Isabel Portugal[1]

[1]Centro de Patogénese Molecular, URIA, Faculdade de Farmácia da Universidade de Lisboa, Portugal; [2]Grupo de Micobactérias, Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (IHMT/UNL), Lisboa, Portugal; [3]Public Health Laboratory: Mycobacteriology/Tuberculosis, Public Health Department, Administração Regional de Saúde de Lisboa e Vale do Tejo, I.P., Lisboa, Portugal; [4]Serviço de Infecciologia, Hospital Curry Cabral, Lisboa, Portugal; [5]Departamento de Doenças Infecciosas, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon, Portugal; [6]Centro de Recursos Microbiológicos (CREM), Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; [7]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London, UK. WC1E 7HT; [8]Pathogen Genomics Laboratory, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia; [9]Sydney Emerging Infections and Biosecurity Institute and School of Public Health, Sydney Medical School, University of Sydney, NSW 2006, Australia

Manuscript in preparation

7.1 ABSTRACT

MDR and XDR-TB present a challenge to TB control and elimination goals. In Lisbon, Portugal, particular and successful XDR-TB strains have been found in circulation for almost two decades.

In the present study we have genotyped and sequenced the genomes of 56 *M. tuberculosis* isolates recovered mostly from Lisbon. While genotyping data revealed three major clusters associated with MDR-TB, two of which associated with XDR-TB, the genomic data contributed to elucidate the phylogenetic positioning of circulating MDR-TB strains, showing a high predominance of a single SNP cluster group 5. Furthermore, a genome-wide phylogeny analysis from these strains, together with 19 publicly available genomes of *M. tuberculosis* clinical isolates, revealed two major clades responsible for M/XDR-TB in the region: Lisboa3 and Q1

The data presented by this study contributes to the expanding knowledge of *M. tuberculosis* genomic diversity yielding insights on microevolution and identification of novel compensatory mutations associated with RIF resistance in *rpoB* and *rpoC*. The screening for other structural variations revealed putative clade-defining variants. One deletion in PPE41, found among Lisboa3 isolates, is proposed to contribute to immune evasion and as a selective advantage. IS mapping has also demonstrated the role of IS*6110* as a major driver in mycobacterial evolution by affecting gene integrity and regulation.

The analysis of non-synonymous/synonymous ratios revealed heterogeneities across the chromosome, genotype and Clusters of Orthologous Groups, highlighting possible and different evolution strategies. Globally, our data supports the notion of a growing genomic diversity facing both setting and host adaptation.

7.2 INTRODUCTION

TB is responsible for approximately 1.4 million deaths each year and is considered a Global Health Emergency by the WHO. Portugal is the Western European country that over the last few decades has had one of the highest TB notification rates in Europe (24.7 cases per 100 000).[21] Although this rate is considered intermediate, the difficulty is the growing threat of drug resistance. In particular, the two most difficult-to-treat forms: MDR-TB (resistance to the two most powerful first-line drugs – INH, and RIF and XDR-TB (MDR plus resistance to a FQ and a second-line injectable drug).[20, 503]

The TB situation in the capital city, Lisbon (incidence 31.5 cases / 100 000 in 2010) has been extensively studied. Laboratory data on resistance prevalence point to high XDR-TB rates in the region, which in recent years have ranged between 44.3-66.1% of the MDR-TB clinical isolates.[482] Genotyping studies using RFLP-IS*6110* and more recently, through the characterization of MIRU-VNTR, have led to the identification of a family of close genetic clusters: the Lisboa family, highly associated with MDR- and XDR-TB.[24, 26, 405] The prevalence of this family in the region may account to up to 74.0% and 80.0% of MDR- and XDR-TB cases, respectively.[24, 25, 482] Another genetically close and endemic cluster, named Q1 also plays an important role in MDR- and XDR-TB in the region and its impact on public health and drug-resistant TB in the region has been addressed in previous publications.[25, 482]

The etiologic agents of TB are the bacterial (sub)species belonging to MTC, such as *M. tuberculosis* or *M. bovis.*[2, 84] *M. tuberculosis* has been regarded for many years as a genetically monomorphic pathogen. Nevertheless, the high-throughput genomic sequencing of diverse clinical strains has revealed a higher degree of variation than initially anticipated. [52, 426, 504, 505] Next-Generation Sequencing (NGS) technology is allowing new insights on the mode of transmission and evolution of the MTC.[428, 430] Furthermore, the ability to compare, at the genomic level, identical strains in different stages of resistance acquisition can also provide new data on the genomic adaptation and compensation to the fixation of resistance-associated mutations in the host's bacilli population. [427, 430]

The genomic determinants of this family of strains are yet to be characterized. In the present study, we have genotyped and sequenced the genomes of 56 *M. tuberculosis* clinical isolates (sourced from the Lisbon Health Region) with the aim of gaining insights into the genomic diversity and microevolution of prevalent MDR- and XDR-TB circulating strains in the Lisbon region.

7.3 RESULTS

Of 56 *M. tuberculosis* isolates studied, 36 (64.3%) were resistant to INH and RIF and were therefore classified as MDR-TB isolates, of which we were able to determine the resistance to second-line drugs for 24 isolates. In total, 10 MDR-TB isolates were also classified as XDR-TB (Table 7.1).

Genotypic analysis

The 24-*loci* MIRU-VNTR genotyping technique grouped the MDR-TB isolates into three major clusters: Lisboa3-A, Lisboa3-B and Q1 (Figure 7.1). Use of the 12-*loci* set groups Lisboa3-A and -B in a single cluster (Lisboa3, data not shown). Only the Lisboa3-B and Q1 clusters were found to be associated with XDR-TB isolates. Eight of the ten XDR-TB isolates belonged to either Lisboa3-B or Q1 cluster, and one of remaining two strains was found to be Q1-related, raising the possibility of ancestral Q1 XDR followed by posterior divergence from this cluster. No XDR-TB isolate was found to belong to Lisboa3-A cluster, indicative that a strain-specific subset is actually responsible for the Lisboa3 associated XDR-TB in the region.

Genomic analysis

The genetic diversity as assessed by MIRU-VNTR can however underestimate the true genomic diversity present in a given setting. All 56 clinical isolates underwent whole genome sequencing (WGS, Illumina HiSeq 2000 technology, paired end 100-150bp). Raw sequencing data were mapped to the H37RV reference, yielding high coverage data for all isolates (mean read depth per position, mean 249.9, range 44 - 1411 fold; mean 99.1% genome covered, range 98.6 - 99.9%) (Table 7.1). The total number of identified SNPs (point mutations differing from H37Rv) ranged between 488-1465 (mean: 928.0, 26.7% in non-coding regions) (Table 7.1). Of the SNPs on coding regions, 58.5% were considered non-synonymous substitutions yielding a mean non-synonymous/synonymous ratio (N_s/S) of 1.41 (Table 1). AG, CT, GA and TC transitions were found to be the most frequent substitution types (Supplementary Figure S7.1), which is reflected by a mean transversion/transition ratio (T_v/T_s) of 0.62. Overall, across the 56 clinical isolates and 19 publicly available reference strains (F11, CDC1551, KZN1435,

KZN4207, KZN605, KZN_R506, KZN_V2475, UT205 RGTB327, RGTB423, CCDC5180, CCDC5079, CTRI-2, BTB05_552, BTB05_559, S96_129, HN878, R1207, and X122), 9419 genome-wide SNPs were identified by mapping to the reference genome of *M. tuberculosis* H37Rv. The number of small insertions and deletions (indels) detected upon read mapping ranged between 15-175 indels per isolate with a size between 1-59 bp (Table 7.1).

Global phylogenetic analysis using WGS

Using WGS data, the 56 clinical isolates and 19 publicly available strains were assigned into established six SCGs and three PGGs.[52, 64] Overall, at least one isolate belonging to each SCG and subgroups was included in the subsequent analysis. Forty-four (78.6%) of the 56 clinical isolates belonged to SCG 5, reflecting the high prevalence of these strains in Lisbon Health Region (Table 7.1).

A phylogenetic tree was inferred from a set of 9419 genome-wide SNPs (Figure 7.2). It reveals that the two main genetic clusters associated with XDR-TB in the region, Q1 and Lisboa3, constitute two genetically close but distinct clades within the SCG 5. The MIRU-VNTR Lisboa3-A cluster was found to form a monophyletic group within the Lisboa3 clade. The MIRU-VNTR Lisboa3-B clade designation was therefore considered as paraphyletic in the light of a genome-wide SNP phylogeny. The sequenced strain closest to the Lisboa3-Q1 clade is *M. tuberculosis* UT205, a virulent Colombian isolate that according with the present phylogeny shares a more recent common ancestor with Q1 strains than with Lisboa3 strains. The KZN strains, responsible for the XDR-TB outbreaks in South Africa, and belonging to the same SCG as Lisboa3 and Q1 strains, are phylogenetically distinctive, forming a monophyletic clade.

Table 7.1 - Isolate characteristics: DST and data derived from WGS including mapping indicators.

Isolate	DST[a]		SCG	PGG	SNPs								INDELs[b]		Mapping Indicators[c]	
	First-Line	Second-Line (SL)			Non-synonymous mutations (N_s)	Synonymous mutations (S)	Total in Coding Regions (T_c)	Total in Non-Coding Regions	Total	N_s/S Ratio	T_v/T_s Ratio	T_c/Total	Total	Size Range	Mean Read Depth	Coverage (%)
ARS10348	IRS	ETH	5	2	410	296	706	243	949	1.3851	0.5865	0.7439	96	1-24	135.18	98.86
ARS11131	IRSP	CAP AMK OFX MOX ETH	5	2	381	286	667	237	904	1.3322	0.6308	0.7378	95	1-24	101.56	98.89
ARS11285	IRS	AMK OFX MOX	4	2	429	302	731	283	1014	1.4205	0.6031	0.7209	114	1-37	159.38	99.83
ARS11463	I	nd	5	2	382	255	637	241	878	1.4980	0.5575	0.7255	76	1-24	52.42	99.47
ARS11661	IS	nd	5	2	378	284	662	234	896	1.3310	0.6443	0.7388	96	1-24	110.16	98.84
ARS12740	IRSP	ETH	5	2	392	287	679	239	918	1.3659	0.6456	0.7397	95	1-24	97.66	98.81
ARS1717	IRP	OFX ETH	6a	3	263	158	421	160	581	1.6646	0.5651	0.7246	67	1-27	68.89	99.76
ARS1760	I	nd	5	2	379	262	641	242	883	1.4466	0.5725	0.7259	83	1-53	69.23	99.40
ARS1900	IRSEP	CAP KAN OFX ETH	5	2	414	301	715	250	965	1.3754	0.6009	0.7409	101	1-24	148.85	98.75
ARS1930	IRSP	na	5	2	389	268	657	228	885	1.4515	0.6158	0.7424	90	1-52	90.43	98.81
ARS2061	IRP	CAP AMK KAN OFX ETH CS PAS	5	2	379	283	662	239	901	1.3392	0.6346	0.7347	89	1-24	80.51	98.83
ARS2202	IRSP	OFX ETH CS	5	2	404	284	688	243	931	1.4225	0.5889	0.7390	91	1-24	79.46	98.99
ARS2573	I	nd	5	2	400	277	677	231	908	1.4440	0.6250	0.7456	97	1-49	82.71	99.03
ARS3649	IRSEP	KAN OFX ETH	5	2	399	282	681	242	923	1.4149	0.5814	0.7378	91	1-24	72.67	98.92
ARS3806	I	nd	5	2	386	283	669	233	902	1.3640	0.6277	0.7417	85	1-24	195.56	98.84
ARS4857	IRP	na	5	2	381	279	660	234	894	1.3656	0.6418	0.7383	92	1-24	91.12	98.86
ARS5858	IREP	OFX	5	2	395	284	679	243	922	1.3908	0.6401	0.7364	102	1-24	142.10	98.98
ARS6483	IRSEP	OFX ETH	5	2	406	303	709	246	955	1.3399	0.5839	0.7424	98	1-24	92.22	98.79
ARS6539	IS	nd	5	2	407	303	710	249	959	1.3432	0.5933	0.7404	63	1-24	273.98	98.87
ARS6559	I	nd	5	2	394	292	686	233	919	1.3493	0.5953	0.7465	86	1-24	63.55	98.76
ARS7496	IS	nd	2	1	589	410	999	405	1404	1.4366	0.6404	0.7115	175	1-39	109.62	99.30
ARS7571	I	nd	5	2	388	265	653	244	897	1.4642	0.5749	0.7280	93	1-24	123.02	99.48
ARS7860	IS	nd	5	2	378	279	657	232	889	1.3548	0.6318	0.7390	86	1-24	72.59	98.60
ARS7884	IRSEP	OFX ETH	5	2	409	300	709	244	953	1.3633	0.5874	0.7440	100	1-24	177.31	98.81
ARS8437	IRSP	CAP ETH	5	2	405	296	701	247	948	1.3682	0.6005	0.7395	92	1-24	179.07	98.78
ARS8600	I	nd	5	2	407	291	698	238	936	1.3986	0.6262	0.7457	74	1-49	254.44	98.97
ARS9427	I	nd	3a	1	632	411	1043	422	1465	1.5377	0.5773	0.7119	152	1-36	44.07	99.82
FF181_97	IRS	na	5	2	417	300	717	249	966	1.3900	0.6185	0.7422	24	1-9	772.29	98.99
FF291_98	IRS	na	5	2	416	294	710	250	960	1.4150	0.6063	0.7396	31	1-18	658.49	98.97
FF359_98	IRS	na	5	2	419	293	712	254	966	1.4300	0.5946	0.7371	31	1-18	836.93	99.01
FF674_96	Susceptible	na	4	2	428	323	751	276	1027	1.3251	0.6047	0.7313	34	1-33	961.69	99.76
HCC1095_10	IRE	nd	3b	2	429	314	743	292	1035	1.3662	0.5577	0.7179	119	1-24	185.45	99.61
HCC1276_11	IRSEP	CAP AMK MOX OFX ETH	5	2	392	281	673	243	916	1.3950	0.6350	0.7347	96	1-24	159.36	98.81

Table 7.1 (cont)

Isolate	DST[a]		SCG	PGG	SNPs								INDELs[b]		Mapping Indicators[c]	
	First-Line	Second-Line (SL)			Non-synonymous mutations (N_s)	Synonymous mutations (S)	Total in Coding Regions (T_c)	Total in Non-Coding Regions	Total	N_s/S Ratio	T_v/T_s Ratio	T_c/Total	Total	Size Range	Mean Read Depth	Coverage (%)
HCC1470_11	IRSEP	CAP AMK KAN OFX MOX ETH CS PAS	5	2	412	295	707	251	958	1.3966	0.5830	0.7380	99	1-24	96.90	98.76
HCC759_09	IR	ETH	2	1	623	414	1037	418	1455	1.5048	0.6102	0.7127	160	1-28	70.73	99.39
HCC916_10	IRSEP	CAP AMK KAN OFX ETH	5	2	373	281	654	238	892	1.3274	0.6044	0.7332	103	1-24	145.85	98.82
HPV105_09	S	nd	5	2	400	281	681	244	925	1.4235	0.6214	0.7362	37	1-22	591.50	98.93
HPV113_08	IRSEP	ETH	6a	3	222	126	348	140	488	1.7619	0.5945	0.7131	62	1-24	167.28	99.94
HPV115_08	IRSEP	CAP AMK KAN OFX ETH	5	2	312	225	537	196	733	1.3867	0.7318	0.7326	76	1-21	179.43	98.82
HPV157_06	IS	nd	5	2	397	287	684	248	932	1.3833	0.5760	0.7339	90	1-24	96.25	98.87
HPV50_09	Susceptible	nd	5	2	397	291	688	238	926	1.3643	0.6554	0.7430	69	1-24	281.12	98.94
HPV51_09	Susceptible	nd	3c	2	420	309	729	276	1005	1.3592	0.5677	0.7254	29	1-18	632.88	99.60
HPV65_08	Susceptible	nd	6a	3	247	160	407	162	569	1.5438	0.5837	0.7153	15	1-18	1149.11	99.57
HPV70_09	Susceptible	nd	5	2	380	277	657	224	881	1.3718	0.6343	0.7457	26	1-45	1410.21	99.65
HVNG1	IRSEP	CAP AMK KAN	5	2	308	225	533	201	734	1.3689	0.7082	0.7262	76	1-24	154.30	98.78
IHMT134_09	IRSP	RFB ETH	5	2	332	228	560	203	763	1.4561	0.6935	0.7339	83	1-24	171.49	98.77
IHMT149_09	IRSEP	RFB CAP AMK OFX MOX ETH	5	2	313	226	539	193	732	1.3850	0.6849	0.7363	75	1-21	168.54	98.81
IHMT194_11	IRSEP	RFB CAP AMK ETH	5	2	382	274	656	236	892	1.3942	0.6318	0.7354	83	1-24	47.61	98.96
IHMT288_95	IRSP	RFB ETH	5	2	415	304	719	247	966	1.3651	0.6230	0.7443	98	1-24	197.48	98.84
IHMT295_08	IRSEP	RFB ETH	2	1	528	337	865	358	1223	1.5668	0.7284	0.7073	127	1-59	181.48	99.18
IHMT308_08	IRP	RFB ETH	5	2	334	228	562	201	763	1.4649	0.7524	0.7366	76	1-21	205.47	98.79
IHMT359_03	R	nd	5	2	406	289	695	247	942	1.4048	0.6276	0.7378	31	1-22	727.91	98.93
IHMT361_08	IREP	CAP ETH	5	2	305	232	537	193	730	1.3147	0.6278	0.7356	70	1-24	191.20	98.80
IHMT69_11	IRSEP	RFB CAP AMK ETH	2	1	609	420	1029	415	1444	1.4500	0.6307	0.7126	169	1-39	202.92	99.36
IHMT80_11	IRSEP	RFB CAP AMK ETH	5	2	380	277	657	236	893	1.3718	0.6318	0.7357	87	1-24	63.98	98.95
IHMT82_09	IRS	RFB CAP ETH	5	2	313	221	534	190	724	1.4163	0.6808	0.7376	81	1-24	121.47	98.69

[a]First-Line: I - isoniazid, R - rifampicin, S - streptomycin, E - ethambutol, P - pyrazinamide; Second-Line: ETH - ethionamide, KAN - kanamycin, AMK - amikacin, OFX - ofloxacin, MOX - moxifloxacin, RFB - rifabutin, PAS - para-amino salicylic acid, CS - cycloserine.

[b]Small INDELs called by SAMtools from mapping to M. tuberculosis H37Rv.

[c]Relative to M. tuberculosis H37Rv

na - not available

nd - not done

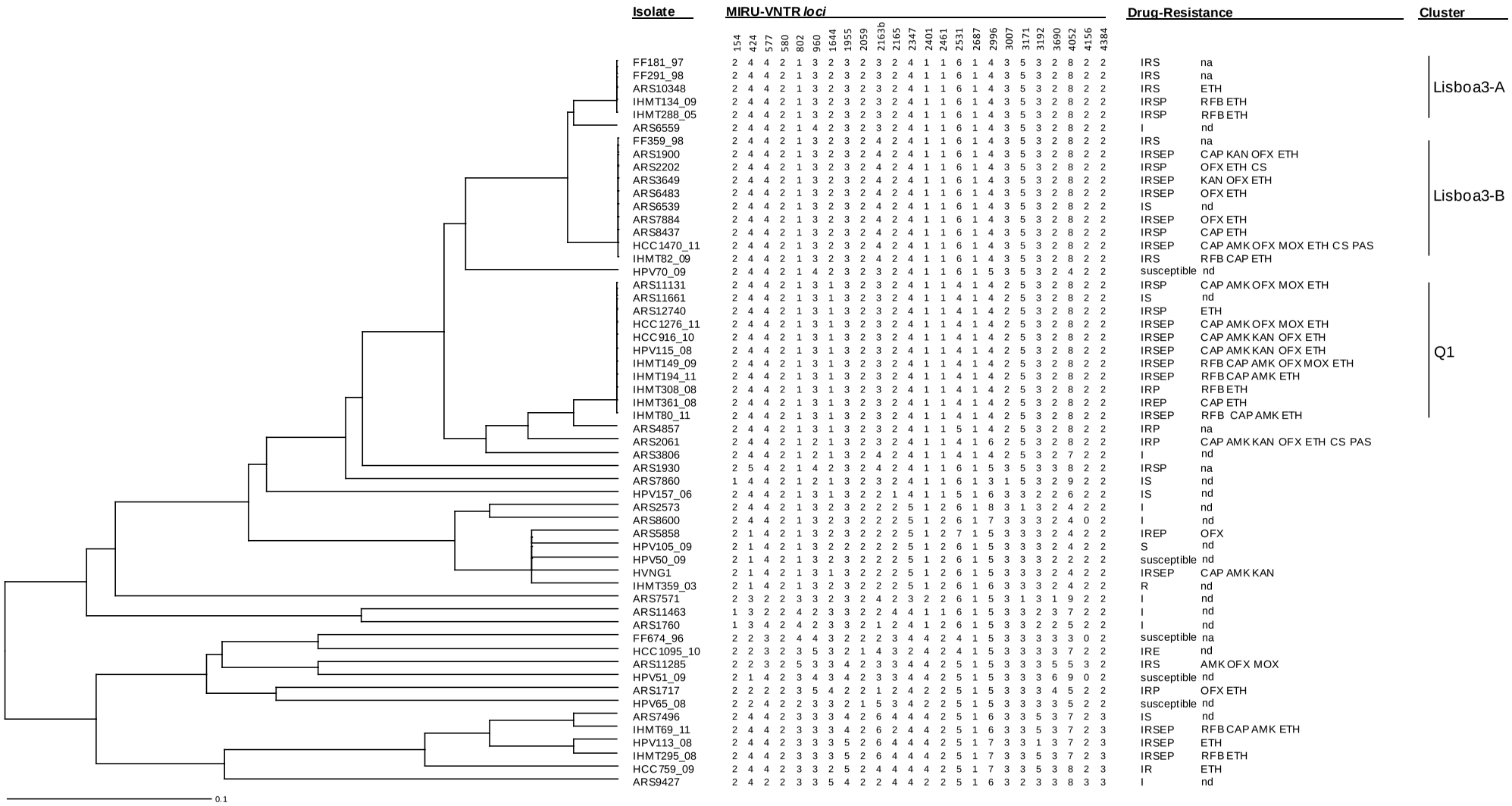


Figure 7.1 – MIRU-VNTR dendrogram of the 56 *M. tuberculosis* clinical isolates subjected to WGS. First-line drug susceptibility testing: I, INH; R, RIF; S, STP; E, EMB; P, PZA. Second-line drug susceptibility testing: KAN, kanamycin; AMK, amikacin; CAP, capreomycin; OFX, ofloxacin; MOX, moxifloxacin; ETH, ethionamide; PAS, para-amino salicylic acid; CS, cycloserine; na, not available; nd, not determined.

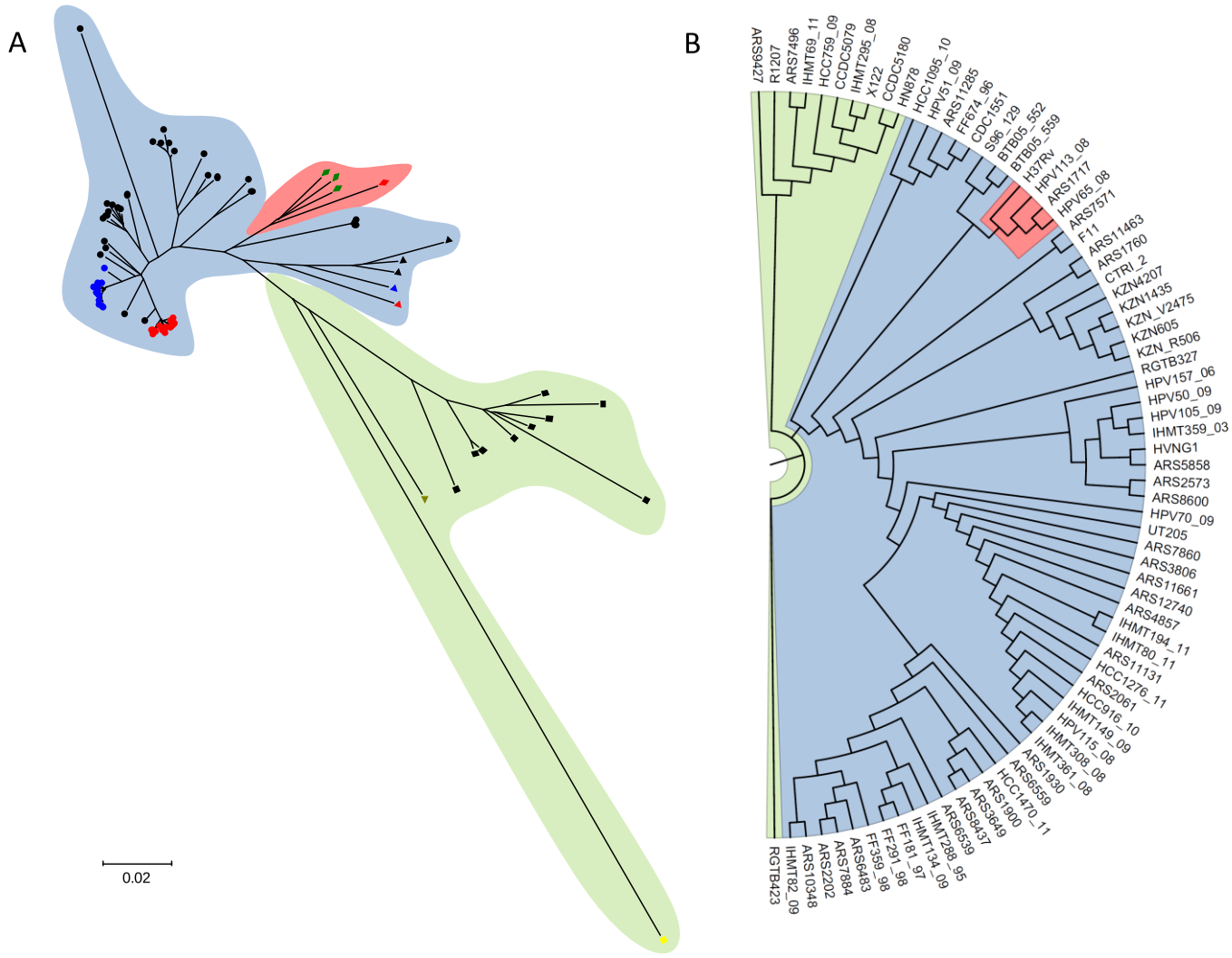


Figure 7.2 – Phylogenetic tree (A) and cladogram (B) of the initial 56 clinical isolates plus 20 *M. tuberculosis* public genomes. PGGs are highlighted in green (PGG1), blue (PGG2) and red (PGG3). A - Isolate-depicting symbols are representative of the different SCGs found in the tree: SCG 1 (yellow square), SCG 2 (black squares), SCG 3a (green triangle), SCG 3b (red triangle), SCG 3c (blue triangle), SCG 4 (black triangles), SCG 5 (circles), SCG 6a (green diamonds), SCG 6b (red diamond). Lisboa3 and Q1 strains are represented by red and blue circles (within SCG 5), respectively.

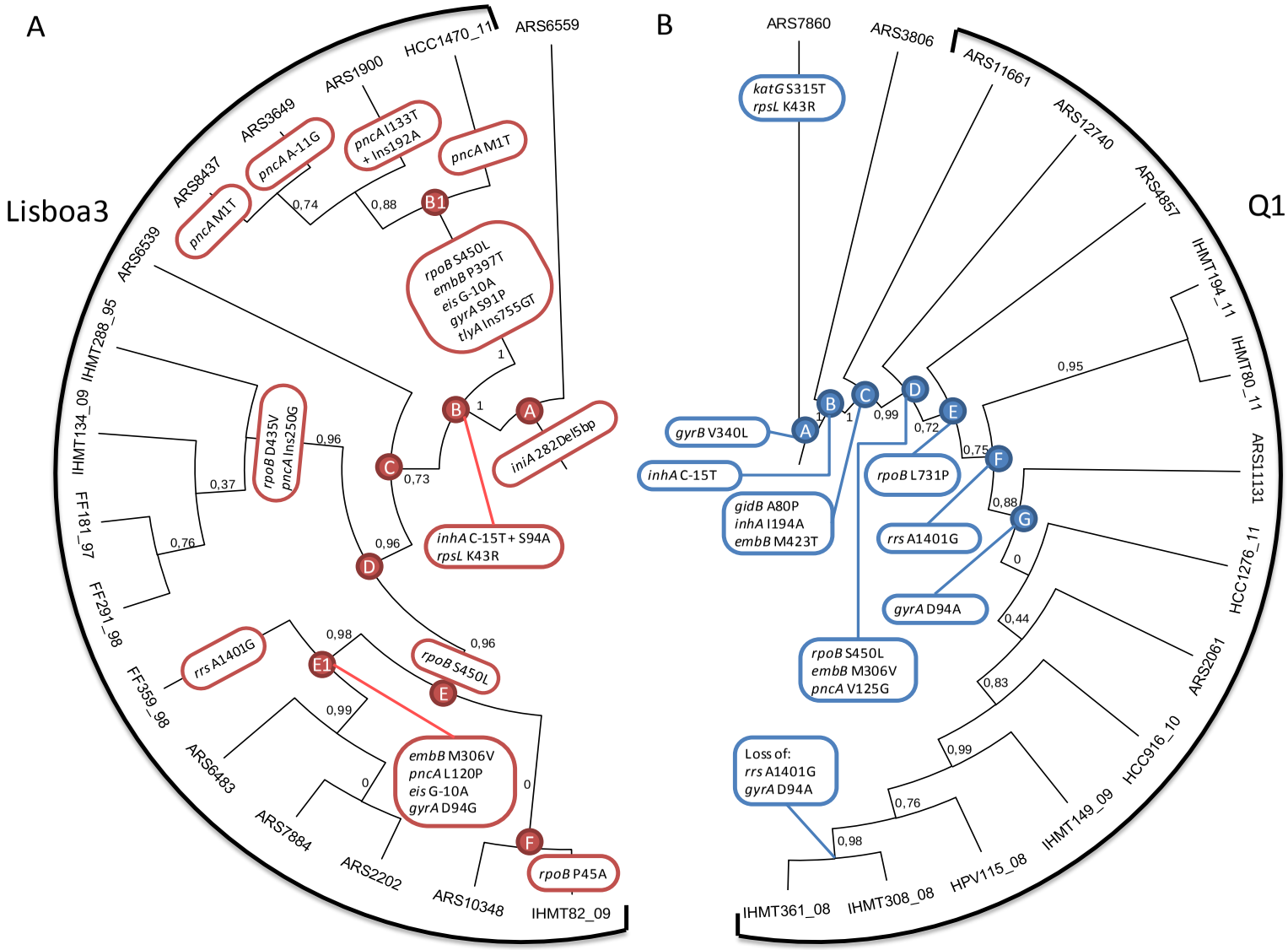


Figure 7.3 –Lisboa3 (A) and Q1 (B) subtree cladograms highlighting the microevolutionary path towards MDR and XDR within these two phylogenetic clades. Mutations acquired in genes associated with first and second-line drug resistance are shown in branch or associated node.

Micro-evolution towards multidrug and extensively drug resistance

Given the relative high number of sequenced strains present in both Lisboa3 and Q1 clades it was possible to trace the microevolutionary path reflecting the genomic changes accompanying the resistance acquisition process. We considered the subtrees containing the Lisboa3 and Q1 clades plus one or two strains for the Lisboa3 and Q1 subtrees, respectively, included as outgroups for the ensuing analysis (Figure 7.3). In particular, we inferred the changes in candidate resistant gene mutations at the nodes of the trees.

The Lisboa3 subtree, including the outgroup strain ARS6559, was found to be characterized by a 5 bp deletion on the *iniA* gene. There is a common acquisition of high-level INH resistance through a *inhA* double mutation (in node B). The data also reflect the acquisition of RIF resistance in three separate occasions, twice in the Lisboa3-B strains by a *rpoB* S450L (equivalent to *E. coli* S531L) and in Lisboa3-A lineage by a *rpoB* D435V (equivalent in *E. coli* to D516V). Acquisition of XDR can be seen in the two branches: the first by acquisition of an *eis* G-10A, *gyrA* S91P and tlyA Ins755GT mutations (node B1); and, by an *eis* G-10A and *gyrA* D94G mutations (node E1). The latter XDR lineage with a *eis* G-10A mutation will most likely present resistance to KAN, but not to CAP and AMK. If drug susceptibility testing to KAN is not included in the standard second-line drug panel of tested drugs, the strains belonging to this lineage will have an undetected XDR phenotype. An exception to this is the strain FF359_98 that bears a *rrs* A1401G mutation that leads to high-level KAN, AMK and CAP resistance.[359] EMB resistance is likely to have been acquired twice by *embB* M306V and P397T mutations. The latter mutation has been previously reported in one EMB resistant isolate.[290] PZA resistance was found to be acquired on multiple independent occasions through *pncA* mutations.

The Q1 subtree included two other Q1-related strains as outgroups and is characterized by a *gyrB* V340L mutation. Here, it is possible to distinguish the acquisition of INH low-level resistance by an *inhA* C-15T mutation (node B) from the acquisition of a higher INH resistance level by an *inhA* missense mutation (I194A, node C).[506] Some of the isolates present in the subtree were found outside the Q1 MIRU-VNTR cluster, but share more recent common ancestors with other strains in the clade, potentially indicating subsequent MIRU-VNTR divergence. The Q1 clade has, therefore, been defined as all isolates bearing the *gidB* A80P mutation characteristic of this cluster and associated with STP intermediate-level resistance previously described by some of us (Chapter 6; Perdigão *et al*, unpublished). A more linear resistance acquisition dynamic was found for this clade. EMB resistance was acquired on two

possible occasions, through an *embB* M423T (node C) and M306V (node D) mutations. RIF resistance development, leading to MDR-TB, was found to be acquired by a *rpoB* S450L mutation (node D), although a second mutation on *rpoB* (L731P) was later developed (node E). Resistance to PZA, injectable second-line drugs and FQs occurred once by mutations on *pncA* (V125G, node D), *rrs* (A1401G, node F) and *gyrA* (D94A, node G), respectively. Interestingly, isolates IHMT308_08 and IHMT361_08 did not show the two latter mutations in *rrs* and *gyrA* genes. Such, is inconsistent with both strains positioning in the Q1 subtree.

A further observation is that M/XDR development in the Lisboa3 subtree appeared to be accompanied by a higher genomic diversification, translated in the number of SNPs and small indels (Supplementary Tables S7.1 and S7.2). Moreover, isolates from the Lisboa3 and Q1 clades were found to bear a mean proportion of 0.73% (range: 0.2-1.8%) and 0.85% (range: 0.2-1.6%) unique SNPs, respectively, in comparison with the total SNP count of each strain. Both clades were found to share a pool of 654 (67.7-90.3%) and 626 (68.2-85.2%) common SNPs, respectively (Supplementary Figure S7.2). This intra-cluster degree of genomic uniqueness is comparable with the data reported by Niemann *et al* for the comparison of two Beijing isolates from the same outbreak clone.[426]

Mutational compensation for RIF-resistance

The acquisition of compensatory mutations following resistance development has been proposed as a possible mechanism to reduce the fitness cost carried by drug resistance.[262] More recently, *rpoA* and *rpoC* genes were found to harbor putative RIF resistance compensatory mutations.[430, 507, 508] The microevolutionary analysis of Lisboa3 and Q1 clades led to the identification of two possible compensatory mutations in *rpoC* (K1152Q, node B to B1 in the Lisboa3 subtree; Supplementary Table S7.3) and *rpoB* (L731P, node D to E in the Q1 subtree; Supplementary Table S7.4) leading to RIF resistance acquisition. The *rpoA* and *rpoC* genes were screened for mutations in all isolates. On the overall 13 different non-synonymous mutations were found, of which only 6 occurred among MDR/RIF-resistance isolates (Table 7.2). The impact on protein function was inferred by computation of SIFT scores.[494] Only three mutations occurring in *rpoC* (Supplementary Figure S7.3) were predicted to affect protein function with SIFT scores equal to 0.00, resulting from the comparison of 189 sequences represented at each position (Table 2). The remaining mutations were predicted to be tolerated and yielded higher SIFT scores (> 0.05), resulting from the comparison of 171-189 sequences representing each position tested (Table 7.2).

We also screened the remaining RNA polymerase subunits, RpoB and RpoZ, but only eight non-synonymous mutations were identified in RpoB, concomitantly with other RIF resistance associated mutations in RpoB (Table 7.2). Five RpoB mutations (P45A, T328N, L452P, L731P and I1106T) were predicted to affect protein function after SIFT score analysis (SIFT score < 0.05) (Table 7.2).

Table 7.2 – Candidate RIF resistance compensatory mutations found in RpoA, RpoB, and RpoC among RIF-resistant isolates with other RIF resistant associated mutations in RpoB

Protein	Mutation	SCG	No. of Isolates	SIFT Score
RpoA	E184D	2	1	0.09
RpoB	P45A	5	1	0.01
RpoB	T328N	2	1	0.03
RpoB	L452P	5	2	0.00
RpoB	V496A	6a	1	0.10
RpoB	D634G	5	1	0.49
RpoB	L731P	2, 5	13	0.00
RpoB	E812G	2	1	0.08
RpoB	I1106T	5	2	0.00
RpoC	G442C	5	1	0.00
RpoC	W484G	2	1	0.00
RpoC	D747G	4	1	0.35
RpoC	K1152Q	5	4	0.00
RpoC	S1287L	6a	1	0.23

Global evolution through Large Sequence Polymorphisms:

Genomes in the *M. tuberculosis* complex can downsize, through LSPs or RDs, and 89 have been previously identified.[3, 78, 81, 509] Across the 75 isolates, 29 (of 89) were detected as absent in at least one isolate (Supplementary Figure S7.4). The most prevalent RDs detected were RD149 (64 isolates), RD152 (45), RD174 (43), RD3 (64), RD6 (54) and RD^{RIO} (43). As expected, all 43 strains bearing the RD174 deletion also had the RD^{RIO} deletion.[510] Both deletions constitute a distinct sub-lineage within the Euro-American lineage and were detected only among SCG 5 strains.[78] UT205, like the Q1 and Lisboa 3 samples, had both deletions, confirming its phylogenetic proximity with these M/XDR associated strains (Supplementary Figure S7.4).

All nine isolates from the SCG 2 had the RD105 deletion characteristic of the East-Asian clade. Of these, other RD deletions were present (RD207 9 isolates, RD181 8, RD142 2). Moreover, other RD deletions associated with specific lineages were detected: RD750 (East-African-Indian

lineage, SCG 3a, 1 isolate), RD115 (Euro-American lineage, Americas-Europe sublineage, SCG 5, 8), RD183 (Euro-American lineage, Americas-Europe sublineage, SCG 3c, 1) RD193 (Euro-American lineage, Americas-Europe sublineage, SCG 4, 3), RD219 (Euro-American lineage, Americas-Europe sublineage, SCG 6a, 3) and RD761 one strain (Euro-American lineage, South Africa sublineage, SCG 5, 1 (F11 strain)) and RD724 (Euro-American lineage, Central Africa sublineage, SCG 5, 3).

No RD region was found to be absent in RGTB423 and only RD^{RIO} deletion was detected in RGTB327. Strain RGTB423 has been found to belong to SCG 1 and PGG 1,[511] but in-silico PCR analysis showed that the strain had the *pks15/1* 7 bp frameshift deletion and the TbD1 deletion indicative of a modern Euro-American strain.[78] Nevertheless, this classification is incongruent with the SCG and PGG classification [84]. On the other hand, RGTB327 was found to have the RD^{RIO} deletion only and *in silico* PCR of the *pks15/1* and TbD1 *loci* also pointed towards a modern Euro-American strain, despite the fact that deletion RD174 was not detected. Further sequencing of these two assembled strains may be required to resolve incongruences.

Structural variability among sequenced strains

Using the sequence data, we attempted to detect structural variants (SVs), including small insertions and deletions (< 100 bp) using the Pindel program and larger variants (≥100 bp) using a combination of different methods compared with local assembly (SVMerge pipeline, Wong *et al*[512]). Among the group of 75 isolates, 2143 different candidate deletions (sizes 1-99 bp) and 4091 different candidate insertions (sizes 1-79 bp) were detected (Supplementary Figures S7.5 and S7.6). We have selected high quality short indels that appear to be clade associated and phylogenetically conserved among all members of the clade (Table 7.3). These variants may be clade-specific and could carry functional consequences that reflect host adaptation and selection.

Table 7.3 - List of selected clade-defining candidate SVs, its position, size and affected ORFs. Each clade-defining candidate SV was selected based on phylogenetic congruence and presence in all members of the specified clade.

Clade	Position[a]	SV Type	Size (bp)	Affected ORF	Function[b]
KZN + CTRI_2	154139	Deletion	53	Repetitive Unit (MIRU)	
Lisboa3	3179243	Deletion	3	Rv2867c	GCN5-related N-acetyltransferase
Lisboa3	560126	Insertion	2	*umaA*	Possible mycolic acid synthase
Lisboa3 subtree (node A)	49479	Deletion	5	Rv0045c	Possible hydrolase
Lisboa3 subtree (node A)	411119	Deletion	5	*iniA*	INH inductible gene protein
Lisboa3 subtree (node A)	922934	Deletion	1	Rv0831c	Conserved protein
Lisboa3 subtree (node A)	1374937	Deletion	1	Rv1232c	Conserved protein; has a magnesium transporter domain (MgtE)
Lisboa3 subtree (node A)	2470887	Deletion	12	Rv2206	Probable conserved transmembrane protein
Lisboa3 subtree (node A)	3302325	Deletion	5	*fadD29*	Fatty-acid-AMP ligase, synthetase, synthase. Involved in biosynthesis of phenolic glycolipids
Lisboa3 subtree (node A)	3560118	Deletion	61	Intergenic	
Lisboa3(node B1)	1472921	Insertion	1	*rrs*	Ribosomal RNA 16S
Lisboa3(node B1)	1656647	Insertion	1	PE_PGRS29	PE-PGRS family protein
Lisboa3(node B1)	1918691	Insertion	2	*tlyA*	2'-O-methyltransferase
Lisboa3-A(node D1)	1769229	Insertion	1	*treY*	Maltooligosyltrehalose synthase TreY
Lisboa3-A(node D1)	2288993	Insertion	1	*pncA*	Pyrazinamidase/nicotinamidase
Q1	3006362	Deletion	1	Rv2689c	Conserved alanine and valine and glycine rich protein
Q1	3652035	Deletion	14	*ctpC*	Probable metal cation-transporting P-type ATPase C
Q1	2009775	Insertion	1	Rv1775	Conserved hypothetical protein
Q1	4408945	Insertion	8	Intergenic	
Q1 subtree (node B)	3956369	Deletion	1	Rv3520c	Possible coenzyme F420-dependent oxidoreductase
Q1 subtree (node B)	4321347	Deletion	1	Intergenic	
Q1 subtree(node B)	3230401	Insertion	7	*glnD*	Probable PII uridylyltransferase
SCG 2	99163	Deletion	9	Rv0090	Possible membrane protein
SCG 2	507029	Deletion	1	Intergenic	
SCG 2	1413130	Deletion	1	Intergenic	
SCG 2	3215623	Deletion	1	Intergenic	
SCG 2	4212833	Deletion	3	Rv3766	Hypothetical protein
SCG 2	55541	Insertion	9	*ponA1*	Penicillin-insensitive transglycosylase N-terminal domain; penicillin-sensitive transpeptidase C-terminal domain
SCG 2 + 3a	4231860	Deletion	89	Rv3785	Hypothetical protein
SCG 2 + 3a	4322040	Deletion	1	Rv3847	Hypothetical protein

Table 7.3 (cont)

Clade	Position[a]	SV Type	Size (bp)	Affected ORF	Function[b]
SCG 2 + 3a	2342650	Insertion	11	Rv2084	Hypothetical protein
SCG 4	4359136	Deletion	18	Rv3879c	ESX-1 secretion-associated protein
SCG 4	4369673	Deletion	1	*mycP2*	Probable alanine and proline rich membrane-anchored mycosin (serine protease)
SCG 4	2081450	Insertion	1	Rv1835c	Conserved hypothetical protein; has two diffrent X-Pro dipeptidyl-peptidase domains
SCG2	3238120	Deletion	3	Rv2923c	Conserved protein; has a Osmotically inducible protein C (OsmC) domain
SCGs 2 + 3a	485811	Deletion	1	*pks6*	Probable membrane bound polyketide synthase
SCGs 3b + 3c + 4	3159999	Deletion	2	Rv2850c	Possible magnesium chelatase
SCGs 3b + 3c + 4	3929090	Deletion	9	*PE_PGRS53*	PE-PGRS family protein
SCGs 3b + 3c + 4	1612279	Insertion	6	Rv1434	Hypothetical protein
SCGs 3b +3c + 4	3462146	Insertion	5	Rv3093c	Hypothetical oxidoreductase
SCGs 3c + 4	1573381	Deletion	1	PE_PGRS25	PE-PGRS family protein
SCGs 3c + 4	1951627	Deletion	1	Rv1725c	Conserved hypothetical protein; has HxlR-like helix-turn-helix domain

[a]Position relative to *M. tuberculosis* H37Rv and corrected from the VCF output from Pindel in order to correspond to the real insertion or deletion sites.

[b]Function extracted from TubercuList (http://tuberculist.epfl.ch/) annotation and domain information from the Pfam database (26.0) (http://pfam.sanger.ac.uk/)

Eight types of larger SVs were detected (Supplementary Table S7.5). Copy number gains were excluded due to a lack of robustness of the methods applied. We have found that putative SVs when detected in more than one isolate mostly present a phylogenetically incongruent distribution. The candidate structural variants found in the present study are reported (Supplementary Figure S7.7), some are identical, but only have similar breakpoints due to uneven coverage. We identified characteristic deletions for the Lisboa3 clade and the ARS6559 isolate (complete Lisboa3 subtree in Figure 4) (112 bp, position 2727803, PPE41 gene), as well as the Q1 clade strains (297 bp, position 3929891, ORF PE_PGRS53). Both deletions were also validated by mapping coverage, but further laboratory confirmation is required.

Insertion sequence mapping and functional consequences for genomic stability

Transposition events from ISs can have a profound effect on strain physiology given the possibility of interference with gene expression by ORF knock-out or gene upregulation resulting from upstream transposition.[88, 513] For all strains included in the phylogenetic analysis, we attempted to map the site of all ISs annotated as mobile elements in the genome of *M. tuberculosis* H37Rv, namely IS*6110*. Some complex inversions were found to be predominantly transpositional events from multi-copy mobile-elements, such as IS*6110*. The analysis revealed the presence of IS*6110*, IS*1081*, IS*1547*, IS*1557* and IS*1558* in multiple copies, but differing in size or annotated sequence at both extremities. For this reason these ISs have been excluded from the mapping analysis.

Variability was only observed for IS*1561* and IS*1532* (Supplementary Figure S7.8). As expected, IS*1561* was not detected in all isolates bearing the RD^{RIO} deletion, whereas IS*1532* is absent in isolates bearing the RD6 deletion found on different SCGs. For IS*6110*, a total of 251 candidate insertion sites have been obtained (Supplementary Figure S9), classified as of high (160), medium (18) or lesser (73) confidence. Almost half (125 (49.8%)) of the 251 ISs were observed on the positive strand. A total of 105 (41.8%) insertion sites were found to be intergenic, from which 64 (25.5%) were in the same orientation with an upstream ORF, known to exert a putative upregulatory effect. For these latter insertion sites the distance from the 3' end to the upstream ORF ranged between 0-939 bp (47 (18.7%) less than 300 bp). Thirty-three sites were found to be within PE/PPE genes, while three other insertion sites were located 18-38 bp upstream of a PPE gene.

Lisboa3 and Q1 clades were found to share 7 IS*6110* sites but were differentiated by IS*6110* insertions on positions 889015 (intergenic) and 4183431 (Rv3732 knock-out) for Lisboa3 and, on 2582457 (intergenic) for Q1 isolates (Figure 7.4). Moreover, we have found that strains belonging to Lisboa3-A MIRU-VNTR cluster (*rpoB* D435V clade on Figure 7.3-A) share three distinct IS6110 insertion sites on Rv1682 (position 1906425), Rv2818c (position 3125900) and Rv3096 (position 3465467). Strains from the XDR-TB Lisboa3 B1 clade (Figure 7.3-A) share a distinctive IS*6110* site on the *plcC* gene (position 2628462). Although no common IS*6110* site was found for the SCG 5 strains, SCG 2 strains were found to share three IS*6110* sites: an intergenic site on position 888786; on Rv1754c (position 1986639); and, on Rv2820c (position 3127931). SCG 4 strains were found to also share three IS*6110* sites on *mmpS1* (position 483580), PPE46 (position 3377326) and PPE47 (position 3379768). One hundred and fifty-three (60.0%) sites were found to be specific to a single isolate.

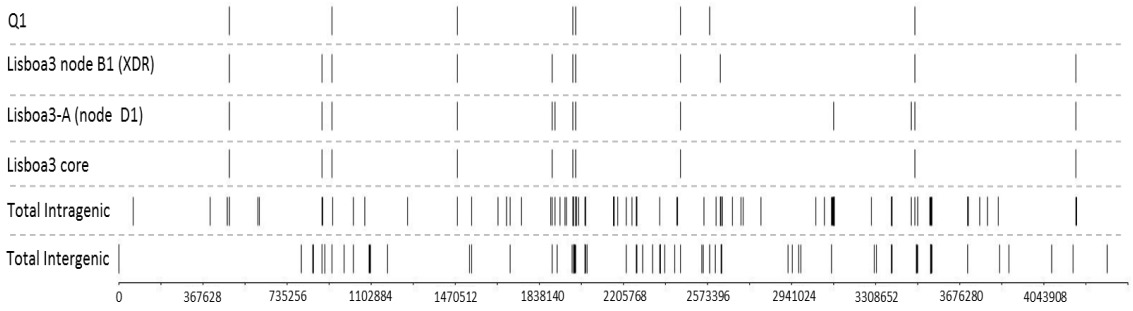


Figure 7.4 – Genomic distribution of total mapped IS6110, intra and intergenic, and insertion sites found among Lisboa3 and Q1 isolates. Lisboa3 core and Q1 lanes depicts all insertion sites that are common to all Lisboa3 and Q1 clade isolates, respectively. Lisboa3 node B1 comprises a XDR-TB lineage shown here with an extra IS6110 copy. Lisboa3-A (node D1) are shown here to bear three additional IS6110 copies when compared with the Lisboa3 core.

Interestingly, an IS*6110* insertion in the NTF locus (position 3493907) was detected in six out of the eight Beijing strains included in the analysis, which is a characteristic of the Beijing/W family (Supplementary Figure S7.9).[514, 515] Hence, two of the three Beijing isolates recovered in Lisbon Health Region were found to belong to the Beijing/W family. No relation with the New York City Beijing/W MDR clade was found as a second insertion in the NTF locus was not detected in any strains.[514, 515] Curiously, a SCG 6a strain (HPV113_08) shared the latter insertion site with the Beijing/W strains, although only one end was detected which can be indicative of another genomic rearrangement. A SCG5 strain (HPV157_06) was found to have an IS*6110* 67 bp upstream of the characteristic IS*6110* insertion site of the Beijing/W family, however in a different orientation. Both insertion sites are found within the Rv3128c. This latter gene has an in-frame amber nonsense mutation in H37Rv and for this reason any functional consequence of IS*6110*-mediated ORF disruption is highly questionable.

Strains belonging to PGG2 were found to have a significantly lower number of IS*6110* copies when compared with PGG1 strains (Kruskal-Wallis test, $p<0.001$). Given the reduced number of PGG3 strains no statistical comparison was possible to perform.

Differential substitution ratios highlight different genomic adaptation strategies

A statistically significant N_s/S ratio was observed between Lisboa3/Q1 and Beijing strains and others, but only the Lisboa3 and Q1 result met a multiple comparison threshold

(Supplementary Table S7.7). The only significant T_v/T_s ratio occurred for differences between Lisboa3 and Q1 clusters (Q1 greater, mean difference: 0.045, p=0.033) (Supplementary Table S7.6).

These ratios were also found to vary across the genome and across the different Clusters of Orthologous gene Groups (COGs). For each strain, we have computed the N_s/S and T_v/T_s ratio for the different genomic quadrants and for each COG. Overall quadrant N_s/S and T_v/T_s comparison, showed that N_s/S ratio varied along the chromosome such that the second quadrant had a lower N_s/S ratio when compared with the other three quadrants and that the first quadrant had the highest N_s/S mean ratio (Kruskal-Wallis, p<0.001) (Supplementary Table S7.7). No statistical difference was observed between the third and fourth quadrant. Regarding the T_v/T_s ratio, an approximately inverse situation was found as no statistical difference was observed between the first, third and fourth quadrants. The second quadrant showed however, a significantly higher T_v/T_s ratio than the three other quadrants (Kruskal-Wallis, p<0.001) (Supplementary Table S7.7).

When these results were stratified by genetic clade, it was found that in the first quadrant the Beijing strains showed a statistically lower N_s/S ratio upon comparison with Q1 and other non-clustered (NC) strains, but not Lisboa3 (Supplementary Table S7.8). No statistical difference was found in this quadrant for T_v/T_s ratio. In the second quadrant, Lisboa3 strains showed a statistically significant reduced N_s/S ratio compared with the other three groups of strains, while Beijing strains presented a higher N_s/S ratio than the remaining groups (Supplementary Table S7.8). Inversely, the T_v/T_s ratio on the second quadrant was significantly higher for Beijing strains when compared to Q1 and other NC strains, but not to Lisboa3 strains (Supplementary Table S7.8). The analysis of the third quadrant showed no statistical difference for Ns/S ratio while Beijing strains showed a higher T_v/T_s ratio on comparison with Lisboa3 and other NC strains, but not Q1 strains. Lisboa3 strains showed a reduced T_v/T_s ratio on this latter quadrant when compared to all other groups. In the fourth quadrant, only a statistical difference was observed for a Q1 reduced N_s/S ratio when comparing with the other strain groups and no significant difference was observed for the T_v/T_s ratio (Supplementary Table S7.8).

These results show that the N_s/S and T_v/T_s ratio measures appear to vary on a strain and chromosome region dependent mode. Data stratification by isolate and quadrant showed that the T_v/T_s ratio was found to correlate negatively with the N_s/S ratio (Pearson, p<0.001).

Correlation between overall isolate N_s/S and T_v/T_s ratio was also attempted but no correlation was found (Pearson, p=0.433).

The comparison of the N_s/S and T_v/T_s ratios across the different COGs also yielded strain dependent results. On comparison with the other three strain groups: Lisboa3 strains showed higher Ns/S ratios on COG groups D (Cell Cycle Control, Mitosis and Meiosis) and P (Inorganic Ion Transport); Q1 strains showed higher Ns/S ratios on COG group V (Defense Mechanisms); and, Beijing strains showed higher N_s/S ratios on COG groups F (Nucleotide Transport and Metabolism), K (Transcription), N (Cell Motility), O (Post translation Modification, Protein turnover and Chaperones) and Q (Secondary Metabolites Biosynthesis, Transport and Catabolism) (Supplementary Table S7.9). Regarding the T_v/T_s ratio no significant difference was observed for Lisboa3 strains, but higher ratios were observed for Q1 strains in COG groups J (Translation), L (Replication, Recombination and Repair), M (Cell Wall, Membrane Biogenesis) and, for Beijing strains in COG group C (Energy Production and Conversion) (Supplementary Table S7.10).

These results support the notion of a differential mode of evolution and adaptation to the human host by accumulation/selection of a higher degree of non-synonymous mutations at genes belonging to specific functional categories.

According to recent work by Namouchi *et al*,[77] the N_s/S ratio varied along the phylogenetic tree, such that terminal branches had a higher N_s/S ratio than inner branches. We have computed the N_s/S and T_v/T_s ratio for the inner nodes assigned in the subtrees in Figure 3 and compared with the respective ratios calculated for the tips of the subtrees. Contrary to the data of Namouchi *et al*,[77] we have verified that both subtrees had ≈6% and ≈12% lower Ns/S ratios at the tips of Lisboa3 and Q1 subtrees, respectively, when compared with the inner nodes of the tree (Independent t-test, $p<0.001$). For the T_v/T_s ratio, the opposite was found: higher T_v/T_s ratios were observed at the tips in comparison with the inner nodes (Mann-Whitney test, $p<0.001$).

7.4 DISCUSSION

For at least two decades the Lisbon Health Region in Portugal has been characterized by a high-level of drug resistance, at first MDR-TB, and later XDR-TB, mainly caused by a particular group of strains: the Lisboa family. Presently, this drug resistance is due almost in its entirety

to an endemic circulation of the Q1 and Lisboa3 phylogenetic clades. Present data from 24-*loci* (not 12-*loci*) MIRU-VNTR allowed the subdivision of the Lisbon3 cluster in two other clusters herein designated as Lisboa3-A and –B. This data suggests two independent outbreaks, over the years, dated back to 90s when the discrimination of Lisboa strains was identified by distinct *rpoB* mutations.[405] The Q1 spoligotyping data has revealed that this cluster is in fact intimately related with the B cluster identified in 90s outbreak (unpublished data). Phylogenetic analysis based on previously published sets of SNPs revealed that Lisboa3 and Q1 strains formed distinct monophyletic evolutionary clades within the SCG 5 and PGG 2.[52, 66] Interestingly, *M. tuberculosis* F11 and the XDR-TB associated KZN strains, both originating from South Africa also belong to SCG 5. Nevertheless a clear distinction is highlighted in the proposed phylogeny. This distinctiveness is also reflected by the RD comparison, but Lisboa3, Q1 and KZN strains appear to have an incongruent phylogeographic association using the RD typing. All these strains belong to the Euro-American lineage according to the RD classification proposed by Gagneux *et al*.[78] However, the KZN strains included in the analysis showed to be positive for RD115, associated with an Americas/Europe sublineage, despite the fact that these strains are a major public health concern in South Africa, namely, the XDR-TB outbreak in KwaZulu Natal.[516, 517] The Lisboa3 and Q1 strains were on the other hand positive for RD174, associated with a West-African sublineage, but constitute a major public health concern in Europe. Present knowledge recognizes that RD174 is also associated with RD^{RIO}, an LSP that has initially been discovered in Rio de Janeiro, Brazil but was later found to be widespread. Historic ties connect Portugal, Brazil and West African Countries and a possible ancestor for these two clades might lie in Africa, more specifically on Portuguese Speaking African Countries. These phylogeographic incongruences are consistent with human migratory events out from, and back into, the African continent.[2]

Another question still seems pertinent as to which selective advantages do these two clades possess allowing such high prevalence in this setting especially since other strains, *e.g.* pre-XDR-TB Beijing strains, also do circulate but at an apparent lesser prevalence? TB caused by RD^{RIO} strains has shown to be associated with weight loss, hemoptysis, higher bacillary loads and progression to cavitary disease.[509, 518] This deletion encompasses several PPE genes that have shown to be a potential source of immune variation (reviewed in Akhter *et al* [519] and Mukhopadhyay *et al* [520]) and hence, may constitute a pathogenic adaptation strategy to immune evasion. Higher bacillary loads are associated with a higher secondary case rate [521-523] and if in fact the absence of these genes truly plays an important role towards an increased virulence, or even transmissibility, it may be a factor that has contributed to the high

prevalence of RD^{RIO} strains in this setting simultaneously contributing to the emergence and spread of M/XDR-TB strains.

The finding that a 112 bp deletion is present among Lisboa3 clade strains, with a more restricted distribution than RD^{RIO}, affecting gene PPE41 might also provide additional clues and contribute to a higher virulence or transmissibility. PPE41 has been previously described has having an immunodominant nature and shown to activate a $CD4^+$ and $CD8^+$ mediated T cell response leading to an enhanced IFN-γ response as well as induce a strong humoral response.[524, 525] The deletion found might constitute a means of immune evasion and constitute a selective advantage over other circulating strains. More specifically, a stronger humoral response to PPE41 was found among extra-pulmonary TB patients.[525] The selective advantage provided by this deletion might therefore also be related with the fact that Lisboa strains were first identified among HIV-infected patients, which is associated with an increase in extra-pulmonary TB.

The use of SNPs as molecular markers has contributed to an improved understanding of the evolutionary history of the *M. tuberculosis* complex. In the present study, given the availability of genomewide SNP data, a SNP-based phylogeny was deduced from the genomic data and, overall, the proposed phylogeny appears to be consistent with other SNP-based phylogenies although as already pointed out: SCG 3 does not exist as a monophyletic lineage but instead as a paraphyletic one. The original report by Filliol *et al* [64] proposed a minimum number of sixteen SNPs that allowed assignment of any strain to an SCG but not to its subgroupings. A later erratum showed that SCG 3a belonged in fact to PGG1 while SCG 3b and 3c belonged to PGG2 as confirmed by our results. Alland *et al* proposed instead a set of nine SNPs that allowed strain assignment to any SCG and each subgroup.[66]

The phylogeny constructed in the present study contributes nevertheless to demonstrate the uniqueness of Lisboa3 and Q1 strains in a global context and will comprise a future framework for genome-wide association studies (GWAS).

The phylogeny proposed also enabled a microevolutionary perspective on the path towards MDR and XDR. As expected, in the Lisboa3 and Q1 clades, INH resistance was found to be mediated by double *inhA* promoter/structural mutations, recently described by some of us to contribute to INH high-level resistance.[506] The acquisition of *inhA* C-15T mutation was found to have occurred independently in both lineages, and in Q1 cluster it was possible to determine that C-15T mutation was acquired at a first stage of INH high-level resistance development. In Lisboa3 it was not possible to determine which mutation appeared in the first place since no

Lisboa3 isolate with a single *inhA* mutation was found. Recent work by Fenner *et al* suggested that *inhA* promoter mutations, more specifically C-15T mutation, might be associated with Lineage 1 (Indo-Oceanic/SCG 1).[84, 526] Nevertheless, another earlier study from Brimacombe *et al* showed that SCG 1 and 5 had all the mutations of interest towards INH resistance.[527] In our view, the fact that INH resistance in both Lisboa3 and Q1 clades is associated with *inhA* mutations, instead of the of the more usual KatG mutations, is possibly related with selective pressures exerted by the drug regimen itself.

The analysis of Lisboa3 subtree has further highlighted the M/XDR evolutive process in this clade. We have recently proposed an evolutionary path regarding drug resistance acquisition dynamics based on the acquisition of an *eis* promoter mutation as the first-step from MDR to XDR.[434] However the SNP phylogeny proposed is consistent with a twice and independent acquisition of an *eis* promoter mutation. Given this phylogeny it is not possible to establish any order of mutation acquisition. Nonetheless, instead of a single event, our analysis supports multiple development of XDR-TB in the same phylogenetic clade. Two different transmission chains involving strains with the RpoB S450L, instead of one, are also more likely since it is proposed that this mutation has also been acquired twice and independently.[405]

One striking phylogenetic incongruence was found in two Q1 strains that lacked both second-line injectable drug and FQ genetic resistance determinants and at the same time sharing a recent common ancestor resistant to these two classes of drugs. These two strains were genotypically and phenotypically susceptible to AMK, CAP and any of the FQs tested. Two explanations may be considered: a phylogenetic misplacement, although the branches had a good statistical support or, these strains may descend from a reverter ancestor. Although theoretically possible, events such as these may be extremely rare. Only one report has documented an in-patient reversion of an isogenic strain from INH resistant to susceptible.[528]

The acquisition of further mutations in *rpoA*, *rpoB* or *rpoC* genes following RIF resistance development was recently demonstrated, using *Salmonella* as a model organism, to have an important role in fitness compensation, leading to a reduction in the doubling-time to values closer to the wild-type.[529] In our microevolutionary analysis we have detected a RpoC mutation (K1152Q) occurring in the same branch as a RpoB S450L (equivalent to S531L in RpoB *E. coli* numbering). It is the first description of a putative compensatory mutation within the Lisboa3 clade, contributing to the success of one of its sub-lineages through the amelioration of the resistance fitness cost.[262] RIF compensatory evolution has been the subject of two recent studies that showed a high prevalence of *rpoA* and *rpoC* mutations mapped to the

RpoA-RpoC interaction region.[507, 508, 530] The *rpoC* mutation described in a Lisboa3 sub-lineage does not fall in this region, nor was it described in these studies.[507, 508] Nevertheless, two other putative compensatory mutations mapping to the RpoA-RpoC interaction region were found in other isolates not belonging to the Lisboa3 or Q1 clades (Supplementary Figure S7.3). The putative role of these two latter mutations is only substantiated by the bioinformatic analysis of residue conservation. However, the putative compensatory role of the Lisboa3 K1152Q RpoC is further substantiated by their co-occurrence in the same branch as the RIF resistance determining mutation in *rpoB*. Furthermore, none of these putative compensatory mutations was previously described and may constitute novel polymorphisms associated with molecular RIF resistance compensation.[430, 507, 508]

RpoB mutational analysis also allowed the identification of five putative compensatory mutations, of which one (L731P) was found to be acquired in the Q1 clade following RIF resistance acquisition through another *rpoB* mutation. This latter mutation was found to be homoplasic as it was also detected in different SCGs, which also points towards the usefulness of this mutation to counteract fitness costs imposed by the acquisition of other RIF resistance associated mutations. Mutations outside the RIF resistance determining region on *rpoB* gene have been described previously on RIF-resistant isolates with no mutations on this region.[249, 261] The mutations herein described as putatively compensatory were only considered as such if a mutation in the RRDR was already present providing further support for the compensatory role of the former.

The role of compensatory mutations in other *loci* and associated with compensation to resistance to other drugs than RIF have been identified and studied, namely, mutations on *ahp*C for INH or on *rrs* for second-line drug aminoglycosides.[531-533] Nevertheless, no compensatory mutations were identified in these genes (data not shown).

Insertion site mapping revealed a high genomic stability of insertion sequences other than IS*6110*. In fact, we have verified that only deletion events were responsible for variability regarding presence/absence of an insertion sequence other than IS*6110*. On the other hand, as demonstrated in this study IS*6110* is a highly polymorphic marker, probably due to its rapid transposition rate.[534, 535]

The finding that 65.5% of the IS*6110* insertion sites mapped were located intragenically is in line with previous reports.[536] Considering that ≈91% of *M. tuberculosis* genome is composed by coding regions,[62] it highlights the deleterious effects of transposition into certain genes essential to viability or to the successful completion of the pathogen's infectious cycle. PGG1

strains, including the Beijing strains, were found to bear a higher number of IS*6110* copies than PGG2 strains. IS*6110* copy number is presumed to be under negative selection,[537] however, in certain circumstances, it is the insertion site *per se* that might provide a selective advantage and not the copy number.

Considering the data obtained in this study, IS*6110* is unarguably the species' most important mobile element when considering transposition impact on genomic integrity. IS*6110* appears to have an important role in genomic re-shaping towards adaptation either through localized disruption of putative antigenic targets (*e.g.* PPE/PE genes) or through its mobile promoter activity located in the IS*6110* 3' end, capable of inducing transcription or upregulating the expression of downstream genes under stressful conditions.[90] We have found a considerable number of insertion sites to be within PPE genes and a more reduced number of sites to be upstream of PPE genes. PPE genes appear to have been positively selected in pathogenic mycobacteria, have important immune and antigenic potential, and some can induce a shift towards a Th2-type response.[519, 520] Not only the IS*6110*-mediated disruption of PPE genes might constitute a mean of immune evasion but it is also conceivable that upregulation of specific PPE genes might affect the Th1/Th2 response balance.

Remarkably, Lisboa3 and Q1 did not show any IS*6110*-mediated disruption of a PPE ORF, nor did we found any IS*6110* upstream of a PPE gene. This fact perhaps demonstrates a different mode of evolution and host adaptation that does not require PPE gene modulation through IS6110 transposition.

Nevertheless, the maximum distance between an IS*6110* and a downstream gene so that this 3' promoter can exert its influence on gene expression is unknown. The results reported by Safi *et al* [90] demonstrate that an IS*6110* in *M. tuberculosis* 210 located 297 bp upstream of Rv1468c was associated with a threefold increase in transcription upon macrophage infection. Our results show that 15% of the mapped sites are located upstream of an ORF in proper orientation and at a distance of less than 300 bp which, at the light of present knowledge, fulfills the necessary assumptions to exert a putative upregulatory effect on those ORFs. Also considering the diversity of ORFs interrupted by IS*6110* copies, gene knock out studies and assessment of downstream gene expression are necessary if a functional role for specific transposition events is to be established.

Spoligotyping lineage association with specific IS*6110* sites has already been demonstrated, highlighting the phylogenetic informativeness of this marker.[538] Our results also support an association of specific IS*6110* sites with strain lineage at both global and local levels.

Notably, the comparison of the distribution of SNPs by COG showed that N_s/S ratios vary through COG in a lineage-dependent manner. Although, Lisboa3 and Q1 isolates might be overrepresented in the analysis due to the high prevalence in the community, we have shown that Lisboa3 and Q1 present statistically different N_s/S ratios from the remaining isolates. We propose that differences in N_s/S COG might highlight different evolution strategies selected during host-pathogen interaction and adaptation.

Moreover, an overall higher N_s/S ratio was observed for the first quadrant and an overall lower N_s/S ratio was observed for the second quadrant revealing heterogeneous N_s/S ratios negatively correlated with the T_v/T_s ratio. The biochemical nature behind this T_v/T_s ratio heterogeneity requires further studies as it may be driving localized higher non-synonymous mutation rates with functional impact on strain evolution. The precise genes affected by non-synonymous mutations within these COG categories merit further studies as each COG includes a considerable number of genes, that mutated might enhance the transmissibility or drug resistance, and should be analyzed in a systems biology perspective using *in silico* models.[539-541]

The finding that terminal branches of the subtrees analyzed had lower N_s/S ratios than the inner branches was contrary to the findings of Namouchi *et al*.[77] Namouchi *et al* suggested that non-synonymous changes might be purged by natural selection yielding higher N_s/S ratios. However, an opposite view is also possible: non-synonymous mutations are favored by natural selection yielding the same higher N_s/S ratio, especially as a mean of adaptation in an organism devoid, or with low, horizontal gene transfer such as *M. tuberculosis.*[72, 76] From our data, we can say that non-synonymous mutations may be favored in the inner branches of both subtrees as a mean to develop and adapt to drug resistance, yielding a higher N_s/S ratio that is consistent with a reduced purifying selection.[2] Nevertheless, the differences between the results from both studies might lie in the fact that the N_s/S ratio analysis herein presented was performed for two sets of strains that are in a much more closer time frame in order to understand microevolution within two clades.

In the present study we have shown that in Lisbon, Portugal, where the MDR-TB situation had already escalated to a XDR-TB situation, it is mainly caused by transmission of two unique phylogenetic clades. We have previously shown that XDR-TB was already a reality in Portugal during the 1990s but noteworthy, the data from the present study clearly shows that these strains belong to the same phylogenetic Lisboa3 M/XDR sublineages that are still presently in

circulation.[434] The uniqueness of these strains was revealed by a distinct phylogenetic placement within SCG 5.

The Lisboa3 clade belongs to a much broader group of strains that usually share at least 90% of MIRU-VNTR pattern similarity: the Lisboa family.[26] One of the future goals is to better understand the populational structure of this family of strains, from which the Lisboa3 clade has differentiated, as strains belonging to this family have previously shown the potential to evolve to MDR-TB.[24]

In conclusion, the molecular diversity described in the present study greatly contributes to the body of knowledge concerning *M. tuberculosis* diversity and supports the notion of higher genomic diversity than the one usually associated with *M. tuberculosis*, mostly acquired through genome downsizing and non-synonymous SNPs. The present study also stresses the need of further genomic studies in order to contribute to a *M. tuberculosis* genome-wide evolutive scenario, representative of different settings. This, together with clinical data, will ultimately enable GWAS with a positive impact in TB management.

7.5 MATERIALS AND METHODS

Isolates and genetic data

The study consists of 56 *M. tuberculosis* clinical isolates (source: 55 Lisbon, 1 Oporto) recovered from laboratories and hospital units across Lisbon Health region. All isolates underwent drug susceptibility testing for INH, RIF, STP, EMB and PZA and second-line drugs using standard methods (see Perdigao *et al* [24]). For MDR-TB strains, drug susceptibility testing was also performed for ETH, CAP, AMK and OFX. DNA extraction was performed from culture growth on Lowenstein-Jensen medium slants using the cetyl trimethylammonium bromide methodology.[445] The DNA was used in genotyping by the 24-*loci* MIRU-VNTR method (see previous work, Supply *et al* [56]). Extracted DNA was also subjected to whole-genome (101bp paired end) sequencing at the KAUST genomics facility using the Illumina HiSeq 2000 platform (500bp insert size). We also complemented this data using sequences in the public domain (F11, CDC1551, KZN1435, KZN4207, KZN605, KZN_R506, CM000789.2; KZN_V2475, UT205 RGTB327, RGTB423, CCDC5180, CCDC5079, CTRI-2, BTB05_552 BTB05_559, S96_129, HN878, R1207, and X122 (all from the NCBI database).

Genomic variant detection

The raw Illumina sequencing data was aligned to the H37Rv reference genome using the Burrows-Wheeler Alignment Tool v.0.6.1.[542] SNPs and small indels (<30 bp) were called using SAMtools software (v0.1.18) to the Variant Call Format (VCF).[543, 544] Other small indels (<100 bp) were detected using the software Pindel.[545] Only variants supported by at least ten sequence reads were considered. Detection of larger structural variants was performed using the SVMerge v1.2 pipeline combining Pindel v0.2.4t, BreakDancer v1.1 Cpp package and, SECluster analysis outputs.[512, 546] Structural variant detection was done for each isolate alone and validation was achieved using comparison with local *de novo* assembly using Velvet.[547] *Loci* reported to be associated with regional differences were identified using the alignments and coverage.[3, 78, 81]

For IS mapping, reads containing specific oligonucleotide sequence of both 5' and 3' extremities (listed in Supplementary Table S7.11) were extracted and, flanking genomic regions of interest concatenated in FASTA format producing one file for each extremity for each strain. Local BLAST analysis (standalone NCBI BLAST v.2.2.27+) was carried out for each file against *M. tuberculosis* H37Rv reference genome, minimum supporting read depth used to as a quality filter (10 for isolates with >500 fold coverage, 2 for the remaining). For IS*6110* BLAST hits, a mapping quality classification scheme was established consisting in high confidence, medium confidence and lesser confidence sites. Paired sites corresponding to mapping of both 5' and 3' ends in all isolates on which it occurred were classified as high confidence sites. Paired insertion sites for which both ends were mapped in at least 50% of the isolates on which they were found to occur were considered of medium confidence. Insertion sites in which only one end of the IS*6110* was mapped were considered of lesser confidence. Furthermore, insertion site hits mapped to *M. tuberculosis* H37Rv were excluded to avoid repetitive mapping.

Other bioinformatics

The genomic data of publicly available *M. tuberculosis* strains (format FASTA) were included in the analysis through conversion to FASTQ format reads using the program dwgsim v.0.1.10, and mapped and analyzed as described above. When necessary, DNA sequence alignment was performed using the CLC Sequence Viewer v7.6.1 (CLC bio™, Aarhus N, Denmark) and visualized in BioEdit v7.1.3.0 (T. Hall).

A MIRU-VNTR-based dendrogram was constructed in the public MIRU-VNTR*plus* database using the D_{sw} measure of genetic distance for tandem repeat *loci* and the UPGMA clustering method.[548] A phylogenetic tree based on SNPs was constructed using Seaview 4.3.5 using the Maximum Likelihood method.[549] The analysis involved 76 nucleotide sequences with a total of 11271 sites in the final dataset. Tree topology was tested using the most recent approximate Likelihood Ratio Test (aLRT) as an alternative to bootstrap.

Putative impact of selected compensatory mutations on protein function was assessed through the use of SIFT scores (available at http://sift.jcvi.org/) computed from the query alignment against UniRef90 database hits (with less than 90% identity, with a median sequence conservation equal to 3.00).[494]

Any statistical analysis was conducted using the SPSS software.

7.6 ACKNOWLEDGEMENTS

This work was partially supported by Project Ref. SDH49: "Early Molecular Detection of M/XDRTB in the Great Lisbon Healthcare Region" from Fundação Calouste Gulbenkian (FCG, Portugal). The sequencing was funded by the KAUST Research Fund. J. Perdigão, D. Machado and C. Silva were supported by FCT grants SFRH/BD/45388/2008, SFRH/BD/65060/2009 and SFRH/BD/73579/2010, respectively.TGC is funded by the Medical Research Council (UK) and Wellcome Trust.

7.7 DISCLOSURE DECLARATION

The authors have no competing interests to declare.

CHAPTER 8

General Discussion

8.1 GENERAL DISCUSSION AND FUTURE PERSPECTIVES

MDR-TB imposes a heavy burden on public health systems and constitutes from a public health perspective a matter of great concern. TB incidence has been declining in Portugal over the last decades, which *per se* is a good indicator.[21, 23] Nevertheless it is MDR-TB and its even more difficult to treat form - XDR-TB, that raises more questions on its management and control. In fact, attending to the global tendency on the number of new cases of MDR-TB and XDR-TB reported by PHA no specific trend can be clearly observed but rather oscillating values.[23]

In the present work we have mainly addressed the study of MDR-TB in the Lisbon Health Region along with its molecular determinants at the epidemiological and resistance levels. One limitation of this work is that all sets of isolates are fundamentally convenience samples obtained through collaborations with the Portuguese National Institute of Health and other laboratories in the region. Despite this, the considerable number of clinical isolates studied, together with the fact that all M/XDR-TB sample sizes are higher than the number of new M/XDR-TB cases reported nationwide should provide some degree of representativity.

This work therefore comprises several studies that were aimed at describing epidemiological aspects of MDR-TB in Lisbon through molecular biology methods and simultaneously, contribute to the global knowledge on *M. tuberculosis* genetic diversity and adaptation to chemotherapy regimens. We started by addressing MDR-TB (Chapter 2) in which we describe the clonality and first-line drug resistance associated mutations in circulating MDR *M. tuberculosis* isolates. We uncovered the main mutations associated with first-line drug resistance. It became noteworthy the high prevalence of *inhA* promoter mutations in INH resistant isolates and not, as usually is expected, a high prevalence of *katG* mutations, namely on the 315 codon. Mutations in the promoter region of *inhA* gene appear to have evolved independently, at least twice, for Q1 and Lisboa3 isolates as demonstrated in Chapter 7. Muller *et al* have described these mutations as a gateway for XDR-TB in South Africa due to its statistical association with XDR-TB.[225] Nevertheless, in Lisbon the INH resistance level is higher than the one anticipated by these authors, due to two different and recently described mutations co-occurring in the *inhA* ORF.[550] For Q1 isolates it was shown in Chapter 7 by microevolutionary analysis that the *inhA* promoter mutation was acquired in the first place while the InhA structural mutation was acquired in a second stage.

Additional mutations are described in Chapter 2 as well as the clonality assessment, which is further discussed below, showing a major prevalence of Lisboa3 strains. However, and perhaps

the most striking observation was in fact the high XDR-TB rate found in the analyzed sample (56%). This high XDR-TB rate motivated the studies on the molecular determinants of resistance to second-line drugs, in Chapter 3, and on the six-year laboratory perspective on the isolation of M/XDR-TB *M. tuberculosis* strains, in Chapter 4.

In Chapter 3, the analysis of 26 XDR-TB clinical isolates showed a high clustering degree with the 26 isolates belonging to one of either two 12-*loci* MIRU-VNTR clusters: Lisboa3 and Q1. The mutational analysis of *gyrA*, *rrs* and *tlyA* genes revealed that cluster-associated mutations were found in all three genes. Worth mentioning, a 2 bp insertion in *tlyA* gene is proposed to be responsible for acquired CAP resistance in Lisboa3 cluster isolates that do not show any classical *rrs* mutation. This finding has strong implications for the implementation in Lisbon of the only commercial NAAT able to screen for second-line drug resistance associated mutations (GenoType™ MTBDR*sl*, Hain Lifescience). The implementation of this test would leave undetected about 58% of the CAP resistant isolates. Moreover, no molecular markers for AMK/KAN resistance were found for these Lisboa3 cluster isolates.

The high rates of XDR-TB and MDR-TB were further investigated by analyzing laboratory records of clinical isolates recovered over a six-year period. In Chapter 4 we have shown that the high MDR-TB and XDR-TB were, on the overall, maintained over the period studied. These observations indicate that the M/XDR-TB high rates are not due to a punctual outbreak but to a continued circulation of M/XDR-TB strains mostly belonging to the previously identified clusters Lisboa3 and Q1. Moreover, an increase in the fraction of MDR-TB resistant to all five first-line drugs was observed. The increase in this fraction over isolates that are resistant to a lesser number of first-line drugs might illustrate a situation in which standardized treatment results in selection of strains with additional resistances leading to resistance amplification. The repetition of this phenomenon allied with transmission can result in more cases of primary drug resistant TB.

It is also important to address here the differences in MIRU-VNTR profiles between the same clusters (Chapters 2, 4, 6 and 7). MIRU-VNTR reproducibility may be affected by difficulties in sizing and allele assignment. The initial report on MIRU-VNTR sizing by capillary electrophoresis used the GeneScan 2500 ROX (Applied Biosystems™) molecular weight marker which is adequate for non-denaturing systems. Later, the MIRU-VNTR technical guide recommended the use of MapMarker™ 1000 ROX.[551] Our experience with both markers show that the use of MapMarker™ 1000 ROX results in a much more accurate peak sizing than with GeneScan™ 2500 ROX. Subsequently, MIRU-VNTR typing and calibration kits have been

developed and commercialized by Genoscreen™ for both improved multiplex amplification and sizing, respectively. Nevertheless, the price of these kits for routine use in molecular epidemiological surveys is still prohibitive in many settings. This latter kit recommends using GeneScan™ 1200 LIZ as a molecular weight marker; however, we prefer MapMarker™ 1000 ROX as electrophoresis visual inspection of raw data is possible given the regular band pattern of the latter. MIRU-VNTR profiles of Lisboa3 and Q1 in Chapter 7 should now be considered the correct ones as sizing was carried out using a custom MapMarker™ molecular weight marker derived from MapMarker™ 1000 ROX but, with additional bands up to 1200 bp. Additionally, a calibration step was also performed through the sizing of MIRU-VNTR *loci* allelic ladders from the Genoscreen™ MIRU-VNTR Calibration Kit with the custom MapMarker™ 1200 ROX.

MIRU-VNTR typing started to be implemented in our laboratory in 2004 for typing of strains recovered in previous years. Peak sizing difficulties that affect the inter- and intralaboratory reproducibility of the technique hamper direct comparison with strains genotyped in the following years (between 2004-2006).[552] For this reason, genotype comparison in Chapter 4 was only made for strains recovered between 2004-2006. Moreover, looking back to our results with the strains from 2003 (Chapter 2) we would now expect a higher clustering rate taking in account the distribution of *pncA* gene mutations and knowledge from microevolution analysis (Chapter 7). Reconstruction of a calibrated MIRU-VNTR allele sizing table is recommended for implementation of MIRU-VNTR typing (Chapter 7).

Given the fact that all the studies in Chapters 2-6 involved genotyping with 12-*loci* MIRU-VNTR one of the main objectives was to progress towards the implementation of a technique with a more powerful discriminative power. To accomplish this, genotyping using additional MIRU-VNTR *loci* was successfully implemented up to the 24-*loci* MIRU-VNTR typing method. Therefore, in Chapter 7 lies also the first proposal for 24-*loci* MIRU-VNTR profiles of Lisboa3 and Q1 strains. The inclusion of further *loci* allowed a better discrimination of these clusters with special importance for the Lisboa3 cluster. It was possible, by using 24 *loci* for MIRU typing, to identify two clusters: Lisboa3-A and –B, of which only Lisboa3-B was found to be associated with XDR-TB. Using a lesser number of *loci* has nevertheless the advantage of identifying more deep phylogenetically related isolates that otherwise may appear as outgroups and not as strains that have undergone MIRU allelic diversification. The ability to now genotype *M. tuberculosis* clinical isolates by the 24-*loci* MIRU-VNTR method will also be used in upcoming studies that will support new perspectives on the populational genetic structure of *M. tuberculosis* in the region (unpublished data). The use of specific models to

analyze this type of multilocus data will enable the definition of the number of populations present within a setting that, if cross-referenced with the patient demographic data, might allow the development of programmatic intervention strategies at the community level.

Since the description of mutations associated with second-line drug resistance that the cross resistance between the second-line injectable drugs has raised questions concerning the mechanisms behind the observed resistance patterns. A study published by Zaunbrecher *et al,* in 2009, revealed that overexpression of the *eis* gene, coding for an acetyltransferase, was responsible for resistance acquisition towards KAN but not AMK.[372] This finding *per se* is interesting in the way that it constitutes a novel drug resistance mechanism in *M. tuberculosis*: overexpression of a drug-modifying enzyme. In Chapter 5 we found that in an extended set of isolates, including those analyzed in Chapter 3, 19 had an *eis* promoter mutation and, proposed a model of stepwise resistance acquisition in Lisboa3 strains. However, the microevolution analysis (Chapter 7) showed that *eis* G-10A mutation is proposed to have been acquired independently in Lisboa3 clade evidencing multiple resistance development and convergent evolution within a cluster. This more recent phylogenetic analysis based on genomewide SNPs is consistent with multiple emergence of XDR.

KAN resistance may therefore be explained by *eis* promoter mutations. It is, in principle, more likely that this *eis* promoter mutation has been selected by a KAN-containing regimen. Athough KAN is said not to be used to treat TB in Portugal as it is not available, the fact that two isolates recovered in the 90s already had the profile of XDR-TB may be consistent with KAN use in TB treatment regimens in the past. It also highlights the fact that XDR-TB transmission was already ongoing in Portugal for at least nine years before the very existence of a XDR-TB definition.[432]

Eis is a known virulence factor that is associated with cellular surface and released to the medium.[489, 553] Wei *et al* have shown that *M. tuberculosis* Eis overexpression in *M. smegmatis* enhances intracellular survival in macrophages while Wu *et al* showed that *eis*-deficient *M. tuberculosis* strains show decreased growth on MonoMac6 cells.[554, 555] Furthermore, it has been suggested that differences in *eis* gene expression levels might translate in different growth rates on MonoMac6 cells by comparison of growth rates and *eis* expression levels in both *M. tuberculosis* H37Rv and TB294 (strain 210 isolate).[555] Upregulation of *eis* gene as a result of selection for KAN resistance may, nevertheless, carry out an increase in virulence. If such is to be confirmed, the use of KAN in treatment regimens should also be revised as it may select for strains with enhanced virulence. Also, this event would contribute to an increased

virulence by certain strains within the Lisboa3 clade and help to explain the high prevalence of these strains in Lisbon Health Region.[24, 482]

One other question remains: despite the fact that *eis* gene product is reported to acetylate KAN with a much higher affinity than AMK, does this lesser affinity result in an AMK MIC increase of clinical significance? Only a two-fold increase in the AMK MIC was observed for one strain, while the other strain did not show any change when compared with *M. tuberculosis* H37Rv. Increase of *eis* gene expression has been shown to be triggered by the upregulation of the principal *M. tuberculosis* sigma factor, SigA, upon macrophage infection.[555, 556] One hypothesis would be that this *in vivo* activation combined with a *eis* promoter hypermorphic mutation would result in a higher resistance level than the one that can be determined *in vitro* using standard DST methods. However, *eis* gene expression from a multicopy plasmid in *M. smegmatis* appears to have no effect on AMK MIC.[554] It will be of future interest to determine the AMK resistance level associated with *eis* promoter mutations using a more sensitive method at lower concentrations, such as the qDST. The study of the survival rates of these strains upon macrophage infection, under different AMK concentrations, might also provide additional data relevant to whether or not use AMK in drug regimens designed to treat infections caused by strains with *eis* promoter mutations.

An additional and relevant implication for regimen design concerns *tlyA* mutations as a major mechanism of CAP resistance in the region. Based on the known possible resistance mechanisms a hierarchy for usage of AMK/KAN or CAP can be suggested. As demonstrated by our data, one of the potential mechanisms leading to CAP resistance are *tlyA* mutations, which do not confer resistance to AMK/KAN. Contrary to the recent WHO guidelines, it may be of interest to choose CAP as the primary second-line injectable drug as it will minimize the risk of AMK/KAN resistance development.[203] Although it is also likely that an *rrs* mutation may be acquired during a CAP-containing regimen, and thus yield resistance to both CAP and AMK/KAN, if a *tlyA* mutation is selected instead, KAN/AMK may still be considered as a choice. AMK usage as the primary second-line injectable drug is more likely to select an *rrs* mutation and derailing the use of the remaining second-line injectable drugs.[203]

Another alternative may lie in recovering VIO to TB treatment. VIO has been used in the past to treat TB but was later replaced by its less toxic analogue, CAP. However, and despite toxicity, VIO may prove useful in the treatment of XDR-TB as the most common mutation that confers resistance to AMK/KAN/CAP (*rrs* A1401G) does not appear to interfere with VIO mode of action.[359] In Lisbon, VIO might prove useful in treatment regimens directed at XDR-TB strains

belonging to the Q1 cluster. VIO is probably not active against Lisboa3 strains bearing the *tlyA* 2bp insertion as its mode of action is dependent of TlyA methyltransferase activity.[365, 375]

Regarding STP resistance, screening for *gidB* mutations (Chapter 6) has lead to the identification of a missense mutation associated with Q1 cluster strains. The *gidB* gene has shown to harbor lineage specific mutations.[327] However, to our knowledge this is the first report of an association of a *gidB* mutation with an endemic cluster associated with XDR-TB. Bioinformatic analysis shows that the mutation takes place in a conserved residue of GidB and is predicted to have an impact on protein function. Our results also support the role of this mutation in the development of an intermediate-level resistance to STP. It is also interesting to observe that no Q1 isolate has ever been detected with a *rpsL* or *rrs* mutations. Also considering that some of these isolates have tested negative for STP resistance it is possible that these have been exposed to STP due to the limited therapeutic options available. The observation that these isolates have not developed any other mutation associated with STP resistance, besides the identified *gidB* mutation, suggests that the A80P mutation in GidB seems to be enough to confer STP resistance in an *in vivo* situation.

According to the recent WHO treatment guidelines, STP is on the drug group 2.[203] The placement of STP in this latter group results from the fact that only one aminoglycoside/tuberactinomycin can be part of a drug regimen. WHO guidelines recommend the choice of AMK/KAN over STP due to the toxicity of the latter.[203] Discordant data has been published by Peloquin *et al* showing that STP was associated with lower ototoxicity when compared with AMK/KAN and, comparable nephrotoxicity with AMK, although higher than KAN.[310] Given the STP distinct resistance mechanism it is therefore possible to use this drug to treat drug resistant TB if STP susceptibility is documented. However, genotypic data regarding the *gidB* gene should be provided as it may have an impact on regimen efficacy. Alternatively, STP qDST may provide phenotypic data on the level of STP that will back a STP-containing regimen.

The accumulated data from genotyping of the Lisboa and Q1 strains points toward the endemicity of MDR-TB in Lisbon Health Region and uniqueness of these strains. The Lisboa and Q1 strains might represent unique examples of mycobacterial adaptation to a specific host population and anti-TB drug therapy. It became therefore necessary to characterize these strains more deeply from a phylogenetic standpoint and gain additional insights of this adaptative process at the genomic level. Moreover, a genome-wide characterization approach might enable the identification of specific strain variants that can provide clues for an

enhanced virulence or be used as specific molecular markers. Thus, for a more deep genomic study, we have, in collaboration with partners with access to NGS technologies, sequenced and analyzed the genomes of 56 *M. tuberculosis* clinical isolates, all except one recovered in Lisbon Health Region. This study, by itself, constitutes one of the largest WGS studies involving *M. tuberculosis* clinical isolates. A genome-wide SNP analysis has enabled the construction of a phylogeny showing that Lisboa3 and Q1 cluster isolates comprise two different monophyletic clades that although distinct are phylogenetically close. The main findings of this genomic study are more thoroughly discussed in Chapter 7 but it is noteworthy the identification of novel putative compensatory mutations for RIF resistance, the detection of clade specific structural variants, some of which with putative impact on strain physiology/host-pathogen interaction, or the uncovering of different substitution rates in a strain-dependent manner.

It is of special importance the identification of candidate structural variants for rapid characterization or detection of specific *M. tuberculosis* genotypes. Such genomic variants may prove especially useful in tracking descendants of original strains associated with M/XDR-TB that due to genetic diversification of traditional epidemiological markers can lead to underestimation of transmission.[557] In fact a 112-3 bp deletion was detected in a PPE gene that appears to be important for immune recognition.[524, 525] This genomic property, although not exclusive of the Lisboa3 clade may add up to other genetic traits that contribute to immune evasion and transmissibility.

The detection of structural variants is limited by the read length of the technology used in this sequencing study. Confirmation by standard PCR and sequencing will be required to confirm those structural variants that may be of interest in future studies. Therefore, structural variants specific to Lisboa3 and Q1 clades will be the subject of already ongoing studies that aim to confirm not only the structural variant itself, but also the precise genomic breakpoints. The confirmation of these variants will allow the development of a molecular assay specifically developed for tracking these strains that if implemented in the field might prove extremely useful in containing these isolates and simultaneously reduce the rates of primary M/XDR-TB in Lisbon.

The microevolution study has produced data whose analysis is still in its beginning but already allowed a glimpse of resistance acquisition dynamic and history in the two major clades. A more thorough analysis will be required to assess the putative impact of the mutations that were predicted to occur along each branch of the analysed subtrees. Lisboa3 isolates have shown a higher degree of diversification and multiple drug resistance acquisition. This

probably reflects an earlier emergence of this clade in the region. As such, upcoming studies will benefit from the inclusion of sequence data from additional Lisboa3 strains that will improve the resolution of these studies. These sort of microevolutionary studies capture the genomic adaptation process that has occurred within the population. It may be however of interest and a future objective to pursue the characterization of isolates recovered from the same patient over time. This latter approach may prove useful for the identification of genomic variants favouring a within-host adaptation process that can be performed, *e.g.*, before or after drug resistance acquisition or over a relapse episode.

Regarding the populational genetic structure, it is interesting to mention the fact that Lisboa3 and Q1 cluster isolates belong to the LAM lineage (unpublished data). More specifically, in the case of Lisboa3 cluster to the LAM1 sub-lineage whereas Q1 cluster, according to the recent rules of spoligotyping clade assignment belongs to LAM4. Analysis of the SITVITWEB database shows that 51% of the strains isolated in Portugal belong to the LAM lineage.

LAM strains are more prevalent in Africa, Central America and South America. In Europe, the more prevalent spoligotype lineages are the T and Haarlem types. Portugal therefore seems to have a significant difference in its populational structure when compared with the remainder European countries. Even comparison with Spain, the only country to which Portugal shares its borders, shows that the LAM prevalence in Spain is only 17.8%, considerably lower than in Portugal. Although these differences in the populational genetic structure support the endemism of TB in Portugal, it also poses questions regarding the phylogeographical origin of these strains.

The process of mycobacterial evolution in acquiring successive drug resistance as been highlighted in the two major clusters. It is now important, but not easy, to withdraw conclusions that ultimately lead to measures that when implemented in the field should prevent the emergence of other XDR-TB clusters. Some of the herein proposed options concerning hierarchical use of second-line injectable drugs might contribute, but also important, is to limit MDR-TB standardized treatment when no information on phenotypic resistance is available. Standardized treatment regimens applied to patients with unknown primary drug resistance can lead to resistance amplification by not providing the number of effective drugs necessary to prevent the emergence of further drug resistance.[558-562] The advantage of a standardized treatment consists in rapidly reducing patient infectiousness avoiding prolonged transmission. Holding the treatment might result in an increase in the number of new cases. Nevertheless, Bonilla *et al* showed that culture conversion was found to

be faster among M/XDR-TB patients with DST results dated before or less than 31 days after treatment initiation.[563] In this scenario molecular diagnosis of resistance might provide timely evidence to drug resistance/susceptibility which may also prove useful when no contacts with TB history are known.[167, 564] It is therefore important to stress the role of standard regimens as an effective way of treating TB infections by susceptible strains but, not regardless of the potential for resistance development/amplification.

In addition, the implementation of routine molecular epidemiology to new cases or at least to drug resistant cases may help to identify and contain the spreading of specific clones. As already mentioned, molecular epidemiology is highly useful when conventional contact tracing is ineffective in revealing patient links due to casual contact and identify places where transmission is taking place. Several studies show that only minor fractions of the cases showing identical genotypes were identified as contacts of the source cases.[397, 565, 566] The additional data provided by genotyping may guide the implementation of additional and adequate control measures.[567-569]

Drug resistance is predicted to carry a fitness cost that has been proposed to limit the spreading of MDR-TB by affecting its transmissibility.[262, 263, 570, 571] Another epidemiological model taking in account the heterogeneous fitness costs predicted that relatively fit MDR strains can eventually outcompete susceptible strains.[572] A more recent model incorporating molecular epidemiology data shows that although TB prevalence can decrease up to a point of local eradication, resulting from improved detection and cure rates, the proportion of drug resistant cases will increase even considering resistance fitness costs.[573] However, occurrence of compensatory evolution has not been taken in account in this latter model and, can have epidemiological consequences.[262, 573-575] Transposing these predictions to the Portuguese setting implies that active case-finding strategies especially directed at finding M/XDR-TB cases are necessary as the implemented strategies, successful in decreasing the overall TB incidence, might result in an increase of the M/XDR-TB fraction.

Despite major advances in decreasing TB incidence in Portugal, to prevent further emergence of MDR-TB clones, multidisciplinary surveillance encompassing laboratory data, traditional epidemiology and molecular epidemiology is necessary. It is important to bear in mind that there is a good reason for the fact that MDR-TB has also been described as *"Ebola with wings"*.[17]

REFERENCES

1. *Captain of death: the story of tuberculosis*: University of Rochester Press, 1997.

2. Hershberg R, Lipatov M, Small PM et al. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. *PLoS Biol* 2008; **6**: e311.

3. Brosch R, Gordon SV, Marmiesse M et al. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proc Natl Acad Sci U S A* 2002; **99**: 3684-9.

4. Rothschild BM, Martin LD, Lev G et al. Mycobacterium tuberculosis complex DNA from an extinct bison dated 17,000 years before the present. *Clin Infect Dis* 2001; **33**: 305-11.

5. Formicola V, Milanesi Q, Scarsini C. Evidence of spinal tuberculosis at the beginning of the fourth millennium BC from Arene Candide cave (Liguria, Italy). *Am J Phys Anthropol* 1987; **72**: 1-6.

6. Donoghue HD, Lee OY, Minnikin DE et al. Tuberculosis in Dr Granville's mummy: a molecular re-examination of the earliest known Egyptian mummy to be scientifically examined and given a medical diagnosis. *Proc Biol Sci* 2010; **277**: 51-6.

7. Donoghue HD, Spigelman M, Greenblatt CL et al. Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. *Lancet Infect Dis* 2004; **4**: 584-92.

8. Nerlich AG, Haas CJ, Zink A et al. Molecular evidence for tuberculosis in an ancient Egyptian mummy. *Lancet* 1997; **350**: 1404.

9. Zink A, Haas CJ, Reischl U et al. Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol* 2001; **50**: 355-66.

10. Crubezy E, Ludes B, Poveda JD et al. Identification of Mycobacterium DNA in an Egyptian Pott's disease of 5,400 years old. *C R Acad Sci III* 1998; **321**: 941-51.

11. Zink AR, Sola C, Reischl U et al. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. *J Clin Microbiol* 2003; **41**: 359-67.

12. Allison MJ, Mendoza D, Pezzia A. Documentation of a case of tuberculosis in Pre-Columbian America. *Am Rev Respir Dis* 1973; **107**: 985-91.

13. Salo WL, Aufderheide AC, Buikstra J et al. Identification of Mycobacterium tuberculosis DNA in a pre-Columbian Peruvian mummy. *Proc Natl Acad Sci U S A* 1994; **91**: 2091-4.

14. Tayles N, Buckley HR. Leprosy and tuberculosis in Iron Age Southeast Asia? *American Journal of Physical Anthropology* 2004; **125**: 239-56.

15. Smith I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev* 2003; **16**: 463-96.

16. [Die Atiologie der Tuberculose. Facsimile of the original contribution by Robert Koch in "Berliner Klinische Wochenschrift" 10 April 1882]. *Fortschr Med* 1982; **100**: 539.

17. *Timebomb: the global epidemic of multi-drug resistant tuberculosis*. New York: McGraw-Hill, 2002.

18. Campbell M. What Tuberculosis did for Modernism: The Influence of a Curative Environment on Modernist Design and Architecture. *Med Hist* 2005; **49**: 463-88.

19. Mond DJ, Khan A. Images in clinical medicine. Lucite-ball plombage. *N Engl J Med* 1994; **330**: 1723.

20. World Health Organization. Global Tuberculosis Control 2012. Geneva. World Health Organization. 2012. WHO Publication No. WHO/HTM/TB/2012.6

21. European Centre for Disease Prevention and Control/WHO Regional Office for Europe. Tuberculosis surveillance and monitoring in Europe 2012. Stockholm. European Centre for Disease Prevention and Control. 2012.

22. Direcção Geral de Saúde. Ponto de Situação Epidemiológica e de Desempenho.Relatório para o Dia Mundial da Tuberculose. Lisboa. Direcção Geral de Saúde, Ministério da Saúde. 2011.

23. Direcção Geral de Saúde. Ponto de Situação Epidemiológica e de Desempenho. Lisboa. Direcção Geral de Saúde, Ministério da Saúde. 2012.

24. Perdigao J, Macedo R, Joao I et al. Multidrug-resistant tuberculosis in Lisbon, Portugal: a molecular epidemiological perspective. *Microb Drug Resist* 2008; **14**: 133-43.

25. Perdigao J, Macedo R, Malaquias A et al. Genetic analysis of extensively drug-resistant Mycobacterium tuberculosis strains in Lisbon, Portugal. *J Antimicrob Chemother* 2010; **65**: 224-7.

26. Portugal I, Covas MJ, Brum L et al. Outbreak of multiple drug-resistant tuberculosis in Lisbon: detection by restriction fragment length polymorphism analysis. *Int J Tuberc Lung Dis* 1999; **3**: 207-13.

27. Madigan MT, Martinko JM, Parker J. Brock Biology of Microorganisms. Upper Saddle River, NJ: Prentice Hall 2000.

28. Magee JG, Ward AC. Mycobacterium. In: W.B. Whitman, M. Goodfellow, P. Kämpfer, H. J. Busse, M.E. Trujillo, W. Ludwig, K. Suzuki, A. Parte, eds. *Bergey's Manual of Systematic Bacteriology*. New York: Springer, 2012; 312-75.

29. Lamont EA, Bannantine JP, Armien A et al. Identification and characterization of a spore-like morphotype in chronically starved Mycobacterium avium subsp. paratuberculosis cultures. *PLoS ONE* 2012; **7**: e30648.

30. Singh B, Ghosh J, Islam NM et al. Growth, cell division and sporulation in mycobacteria. *Antonie Van Leeuwenhoek* 2010; **98**: 165-77.

31. Traag BA, Driks A, Stragier P et al. Do mycobacteria produce endospores? *Proc Natl Acad Sci U S A* 2010; **107**: 878-81.

32. Ghosh J, Larsson P, Singh B et al. Sporulation in mycobacteria. *Proc Natl Acad Sci U S A* 2009; **106**: 10781-6.

33. Kaur D, Guerin ME, Skovierova H et al. Biogenesis of the cell wall and other glycoconjugates of Mycobacterium tuberculosis. *Adv Appl Microbiol* 2009; **69**: 23-78.

34. Draper P, Daffé M. The Cell Envelope of Mycobacterium tuberculosis with Special Reference to the Capsule and Outer Permeability Barrier. In: S.T. Cole, K.D. Eisenach, W. R. Jacobs Jr, D. N. McMurray, eds. *Tuberculosis and the Tubercle Bacillus*. Washington, DC: ASM Press, 2005; 261-74.

35. Fisher CA, Barksdale L. Cytochemical reactions of human leprosy bacilli and mycobacteria: ultrastructural implications. *J Bacteriol* 1973; **113**: 1389-99.

36. Klegerman ME, Devadoss PO, Garrido JL et al. Chemical and ultrastructural investigations of Mycobacterium bovis BCG: implications for the molecular structure of the mycobacterial cell envelope. *FEMS Immunol Med Microbiol* 1996; **15**: 213-22.

37. Euzeby JP. List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *Int J Syst Bacteriol* 1997; **47**: 590-2.

38. Euzéby JP. List of Prokaryotic names with Standing in Nomenclature. www.bacterio.cict.fr/ (06/08 2012, date last accessed).

39. Tortoli E. Impact of genotypic studies on mycobacterial taxonomy: the new mycobacteria of the 1990s. *Clin Microbiol Rev* 2003; **16**: 319-54.

40. Tortoli E. The new mycobacteria: an update. *FEMS Immunol Med Microbiol* 2006; **48**: 159-78.

41. Runyon EH. Anonymous mycobacteria in pulmonary disease. *Med Clin North Am* 1959; **43**: 273-90.

42. Stahl DA, Urbance JW. The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria. *J Bacteriol* 1990; **172**: 116-24.

43. Devulder G, Perouse de Montclos M, Flandrois JP. A multigene approach to phylogenetic analysis using the genus Mycobacterium as a model. *Int J Syst Evol Microbiol* 2005; **55**: 293-302.

44. Tortoli E. Phylogeny of the genus Mycobacterium: many doubts, few certainties. *Infect Genet Evol* 2012; **12**: 827-31.

45. Mignard S, Flandrois JP. A seven-gene, multilocus, genus-wide approach to the phylogeny of mycobacteria using supertrees. *Int J Syst Evol Microbiol* 2008; **58**: 1432-41.

46. Cousins DV, Bastida R, Cataldi A et al. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. *Int J Syst Evol Microbiol* 2003; **53**: 1305-14.

47. Aranaz A, Cousins D, Mateos A et al. Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov., sp. nov. *Int J Syst Evol Microbiol* 2003; **53**: 1785-9.

48. Aranaz A, Liebana E, Gomez-Mampaso E et al. Mycobacterium tuberculosis subsp. caprae subsp. nov.: a taxonomic study of a new member of the Mycobacterium tuberculosis complex isolated from goats in Spain. *Int J Syst Bacteriol* 1999; **49 Pt 3**: 1263-73.

49. Niemann S, Richter E, Rusch-Gerdes S. Biochemical and genetic evidence for the transfer of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to the species Mycobacterium bovis Karlson and Lessel 1970 (approved lists 1980) as Mycobacterium bovis subsp. caprae comb. nov. *Int J Syst Evol Microbiol* 2002; **52**: 433-6.

50. Alexander KA, Laver PN, Michel AL et al. Novel Mycobacterium tuberculosis complex pathogen, M. mungi. *Emerg Infect Dis* 2010; **16**: 1296-9.

51. van Ingen J, Rahim Z, Mulder A et al. Characterization of Mycobacterium orygis as M. tuberculosis complex subspecies. *Emerg Infect Dis* 2012; **18**: 653-5.

52. Sreevatsan S, Pan X, Stockbauer KE et al. Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci U S A* 1997; **94**: 9869-74.

53. Gutierrez MC, Brisse S, Brosch R et al. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. *PLoS Pathog* 2005; **1**: e5.

54. Blackwood KS, Wolfe JN, Kabani AM. Application of mycobacterial interspersed repetitive unit typing to Manitoba tuberculosis cases: can restriction fragment length polymorphism be forgotten? *J Clin Microbiol* 2004; **42**: 5001-6.

55. Brudey K, Driscoll JR, Rigouts L et al. Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 2006; **6**: 23-9.

56. Supply P, Allix C, Lesjean S et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. *J Clin Microbiol* 2006; **44**: 4498-510.

57. Supply P, Lesjean S, Savine E et al. Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. *J Clin Microbiol* 2001; **39**: 3563-71.

58. van Soolingen D, Hermans PW, de Haas PE et al. Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991; **29**: 2578-86.

59. Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med* 2003; **349**: 1149-56.

60. Demay C, Liens B, Burguiere T et al. SITVITWEB--a publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. *Infect Genet Evol* 2012; **12**: 755-66.

61. Steenken W, Oatway WH, Petroff SA. Biological Studies of the Tubercle Bacillus : Iii. Dissociation and Pathogenicity of the R and S Variants of the Human Tubercle Bacillus (H(37)). *J Exp Med* 1934; **60**: 515-40.

62. Cole ST, Brosch R, Parkhill J et al. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 1998; **393**: 537-44.

63. Baker L, Brown T, Maiden MC et al. Silent nucleotide polymorphisms and a phylogeny for Mycobacterium tuberculosis. *Emerg Infect Dis* 2004; **10**: 1568-77.

64. Filliol I, Motiwala AS, Cavatore M et al. Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* 2006; **188**: 759-72.

65. Gutacker MM, Mathema B, Soini H et al. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. *J Infect Dis* 2006; **193**: 121-8.

66. Alland D, Lacher DW, Hazbon MH et al. Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of Mycobacterium tuberculosis and the utility of LSPs in phylogenetic analysis. *J Clin Microbiol* 2007; **45**: 39-46.

67. Bifani PJ, Mathema B, Kurepina NE et al. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. *Trends Microbiol* 2002; **10**: 45-52.

68. Blazquez J, Espinosa de Los Monteros LE, Samper S et al. Genetic characterization of multidrug-resistant Mycobacterium bovis strains from a hospital outbreak involving human immunodeficiency virus-positive patients. *J Clin Microbiol* 1997; **35**: 1390-3.

69. Rivero A, Marquez M, Santos J et al. High rate of tuberculosis reinfection during a nosocomial outbreak of multidrug-resistant tuberculosis caused by Mycobacterium bovis strain B. *Clin Infect Dis* 2001; **32**: 159-61.

70. van Soolingen D, Qian L, de Haas PE et al. Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. *J Clin Microbiol* 1995; **33**: 3234-8.

71. Malik AN, Godfrey-Faussett P. Effects of genetic variability of Mycobacterium tuberculosis strains on the presentation of disease. *Lancet Infect Dis* 2005; **5**: 174-83.

72. Jang J, Becq J, Gicquel B et al. Horizontally acquired genomic islands in the tubercle bacilli. *Trends Microbiol* 2008; **16**: 303-8.

73. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L et al. Horizontal transfer of a virulence operon to the ancestor of Mycobacterium tuberculosis. *Mol Biol Evol* 2006; **23**: 1129-35.

74. Becq J, Gutierrez MC, Rosas-Magallanes V et al. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* 2007; **24**: 1861-71.

75. Sassetti CM, Rubin EJ. Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* 2003; **100**: 12989-94.

76. Hirsh AE, Tsolaki AG, DeRiemer K et al. Stable association between strains of Mycobacterium tuberculosis and their human host populations. *Proc Natl Acad Sci U S A* 2004; **101**: 4871-6.

77. Namouchi A, Didelot X, Schock U et al. After the bottleneck: Genome-wide diversification of the Mycobacterium tuberculosis complex by mutation, recombination, and natural selection. *Genome research* 2012; **22**: 721-34.

78. Gagneux S, DeRiemer K, Van T et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2006; **103**: 2869-73.

79. Goguet de la Salmoniere YO, Kim CC, Tsolaki AG et al. High-throughput method for detecting genomic-deletion polymorphisms. *J Clin Microbiol* 2004; **42**: 2913-8.

80. Tsolaki AG, Gagneux S, Pym AS et al. Genomic Deletions Classify the Beijing/W Strains as a Distinct Genetic Lineage of Mycobacterium tuberculosis. 2005; **43**: 3185.

81. Tsolaki AG, Hirsh AE, DeRiemer K et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. *Proc Natl Acad Sci U S A* 2004; **101**: 4865-70.

82. Huard RC, Lazzarini LC, Butler WR et al. PCR-based method to differentiate the subspecies of the Mycobacterium tuberculosis complex on the basis of genomic deletions. *J Clin Microbiol* 2003; **41**: 1637-50.

83. Hill V, Zozio T, Sadikalay S et al. MLVA based classification of Mycobacterium tuberculosis complex lineages for a robust phylogeographic snapshot of its worldwide molecular diversity. *PLoS ONE* 2012; **7**: e41991.

84. Gagneux S, Small PM. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infect Dis* 2007; **7**: 328-37.

85. Newton SM, Smith RJ, Wilkinson KA et al. A deletion defining a common Asian lineage of Mycobacterium tuberculosis associates with immune subversion. *Proc Natl Acad Sci U S A* 2006; **103**: 15594-8.

86. Dos Vultos T, Mestre O, Rauzier J et al. Evolution and diversity of clonal bacteria: the paradigm of Mycobacterium tuberculosis. *PLoS One* 2008; **3**: e1538.

87. Sampson SL, Richardson M, Van Helden PD et al. IS6110-mediated deletion polymorphism in isogenic strains of Mycobacterium tuberculosis. *J Clin Microbiol* 2004; **42**: 895-8.

88. Soto CY, Menendez MC, Perez E et al. IS6110 mediates increased transcription of the phoP virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. *J Clin Microbiol* 2004; **42**: 212-9.

89. Otal I, Gomez AB, Kremer K et al. Mapping of IS6110 insertion sites in Mycobacterium bovis isolates in relation to adaptation from the animal to human host. *Vet Microbiol* 2008; **129**: 333-41.

90. Safi H, Barnes PF, Lakey DL et al. IS6110 functions as a mobile, monocyte-activated promoter in Mycobacterium tuberculosis. *Mol Microbiol* 2004; **52**: 999-1012.

91. Fleischmann RD, Alland D, Eisen JA et al. Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. *J Bacteriol* 2002; **184**: 5479-90.

92. Caws M, Thwaites G, Dunstan S et al. The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. *PLoS Pathog* 2008; **4**: e1000034.

93. Reed MB, Domenech P, Manca C et al. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 2004; **431**: 84-7.

94. Sinsimer D, Huet G, Manca C et al. The phenolic glycolipid of Mycobacterium tuberculosis differentially modulates the early host cytokine response but does not in itself confer hypervirulence. *Infect Immun* 2008; **76**: 3027-36.

95. Manca C, Tsenova L, Barry CE, 3rd et al. Mycobacterium tuberculosis CDC1551 induces a more vigorous host response in vivo and in vitro, but is not more virulent than other clinical isolates. *J Immunol* 1999; **162**: 6740-6.

96. Nahid P, Bliven EE, Kim EY et al. Influence of M. tuberculosis lineage variability within a clinical trial for pulmonary tuberculosis. *PLoS ONE* 2010; **5**: e10753.

97. Lopez B, Aguilar D, Orozco H et al. A marked difference in pathogenesis and immune response induced by different Mycobacterium tuberculosis genotypes. *Clin Exp Immunol* 2003; **133**: 30-7.

98. Coscolla M, Gagneux S. Does M. tuberculosis genomic diversity explain disease diversity? *Drug Discov Today Dis Mech* 2010; **7**: e43-e59.

99. *Mycobacteria and Human Disease, Second Edition*. London: Arnold, 1996.

100. Ormerod LP. Respiratory Tuberculosis. In: P.D.O. Davies, ed. *Clinical Tuberculosis, Second Edition*. London: Chapman and Hall Medical, 1998.

101. Abalain ML, Petsaris O, Hery-Arnaud G et al. Fatal congenital tuberculosis due to a Beijing strain in a premature neonate. *J Med Microbiol* 2010; **59**: 733-5.

102. Escamilla Ocanas CE, Bolanos Jimenez R, Marcos Abdala HG et al. [Description of a case presented in the previous issue: congenital tuberculosis]. *Arch Argent Pediatr* 2012; **110**: 174-5.

103. Stahelin-Massik J, Carrel T, Duppenthaler A et al. Congenital tuberculosis in a premature infant. *Swiss Med Wkly* 2002; **132**: 598-602.

104. Grosset J. Mycobacterium tuberculosis in the extracellular compartment: an underestimated adversary. *Antimicrob Agents Chemother* 2003; **47**: 833-6.

105. Russell DG, Barry CE, 3rd, Flynn JL. Tuberculosis: what we don't know can, and does, hurt us. *Science (New York, NY* 2010; **328**: 852-6.

106. Schluger NW, Rom WN. The host immune response to tuberculosis. *Am J Respir Crit Care Med* 1998; **157**: 679-91.

107. Torrelles JB, Schlesinger LS. Diversity in Mycobacterium tuberculosis mannosylated cell wall determinants impacts adaptation to the host. *Tuberculosis (Edinburgh, Scotland)* 2010; **90**: 84-93.

108. Vergne I, Chua J, Deretic V. Tuberculosis toxin blocking phagosome maturation inhibits a novel Ca2+/calmodulin-PI3K hVPS34 cascade. *J Exp Med* 2003; **198**: 653-9.

109. Fratti RA, Chua J, Vergne I et al. Mycobacterium tuberculosis glycosylated phosphatidylinositol causes phagosome maturation arrest. *Proc Natl Acad Sci U S A* 2003; **100**: 5437-42.

110. Mortellaro A, Robinson L, Ricciardi-Castagnoli P. Spotlight on Mycobacteria and dendritic cells: will novel targets to fight tuberculosis emerge? *EMBO Mol Med* 2009; **1**: 19-29.

111. Raja A. Immunology of tuberculosis. *Indian J Med Res* 2004; **120**: 213-32.

112. Behar SM, Martin CJ, Booty MG et al. Apoptosis is an innate defense function of macrophages against Mycobacterium tuberculosis. *Mucosal Immunol* 2011; **4**: 279-87.

113. Chen M, Gan H, Remold HG. A mechanism of virulence: virulent Mycobacterium tuberculosis strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis. *J Immunol* 2006; **176**: 3707-16.

114. Luna JA. A Tuberculosis Guide for Specialist Physicians. *International Union Against Tuberculosis and Lung Disease, Paris* 2004; **49**.

115. Kim MJ, Wainwright HC, Locketz M et al. Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO Mol Med* 2010; **2**: 258-74.

116. Paige C, Bishai WR. Penitentiary or penthouse condo: the tuberculous granuloma from the microbe's point of view. *Cell Microbiol* 2010; **12**: 301-9.

117. Silva Miranda M, Breiman A, Allain S et al. The tuberculous granuloma: an unsuccessful host defence mechanism providing a safety shelter for the bacteria? *Clin Dev Immunol* 2012; **2012**: 139127.

118. Lillebaek T, Dirksen A, Baess I et al. Molecular evidence of endogenous reactivation of Mycobacterium tuberculosis after 33 years of latent infection. *J Infect Dis* 2002; **185**: 401-4.

119. Singh V, Jamwal S, Jain R et al. Mycobacterium tuberculosis-driven targeted recalibration of macrophage lipid homeostasis promotes the foamy phenotype. *Cell Host Microbe* 2012; **12**: 669-81.

120. Boon C, Dick T. How Mycobacterium tuberculosis goes to sleep: the dormancy survival regulator DosR a decade later. *Future Microbiol* 2012; **7**: 513-8.

121. Shiloh MU, Manzanillo P, Cox JS. Mycobacterium tuberculosis senses host-derived carbon monoxide during macrophage infection. *Cell Host Microbe* 2008; **3**: 323-30.

122. Lim A, Eleuterio M, Hutter B et al. Oxygen depletion-induced dormancy in Mycobacterium bovis BCG. *J Bacteriol* 1999; **181**: 2252-6.

123. Boon C, Li R, Qi R et al. Proteins of Mycobacterium bovis BCG induced in the Wayne dormancy model. *J Bacteriol* 2001; **183**: 2672-6.

124. Voskuil MI, Schnappinger D, Visconti KC et al. Inhibition of respiration by nitric oxide induces a Mycobacterium tuberculosis dormancy program. *J Exp Med* 2003; **198**: 705-13.

125. Kumar A, Deshane JS, Crossman DK et al. Heme oxygenase-1-derived carbon monoxide induces the Mycobacterium tuberculosis dormancy regulon. *J Biol Chem* 2008; **283**: 18032-9.

126. Via LE, Lin PL, Ray SM et al. Tuberculous granulomas are hypoxic in guinea pigs, rabbits, and nonhuman primates. *Infect Immun* 2008; **76**: 2333-40.

127. World Health Organization. Global HIV/AIDS response: epidemic update and health sector progress towards universal access: progress report 2011. Geneva. World Health Organization. 2011.

128. Aaron L, Saadoun D, Calatroni I et al. Tuberculosis in HIV-infected patients: a comprehensive review. *Clin Microbiol Infect* 2004; **10**: 388-98.

129. Pawlowski A, Jansson M, Skold M et al. Tuberculosis and HIV co-infection. *PLoS Pathog* 2012; **8**: e1002464.

130. Rosas-Taraco AG, Arce-Mendoza AY, Caballero-Olin G et al. Mycobacterium tuberculosis upregulates coreceptors CCR5 and CXCR4 while HIV modulates CD14 favoring concurrent infection. *AIDS Res Hum Retroviruses* 2006; **22**: 45-51.

131. Spear GT, Kessler HA, Rothberg L et al. Decreased oxidative burst activity of monocytes from asymptomatic HIV-infected individuals. *Clin Immunol Immunopathol* 1990; **54**: 184-91.

132. Goletti D, Weissman D, Jackson RW et al. Effect of Mycobacterium tuberculosis on HIV replication. Role of immune activation. *J Immunol* 1996; **157**: 1271-8.

133. Shattock RJ, Friedland JS, Griffin GE. Modulation of HIV transcription in and release from human monocytic cells following phagocytosis of Mycobacterium tuberculosis. *Res Virol* 1993; **144**: 7-12.

134. Ku NS, Oh JO, Shin SY et al. Effects of tuberculosis on the kinetics of CD4+ T cell count among HIV-infected patients who initiated antiretroviral therapy early after tuberculosis treatment. *AIDS Res Hum Retroviruses* 2012.

135. Bishop PJ, Neumann G. The history of the Ziehl-Neelsen stain. *Tubercle* 1970; **51**: 196-206.

136. Nolte FS, Metchock B. Mycobacterium. In: P.R. Murray, ed. *Manual of Clinical Microbiology*. Washington, D.C.: ASM Press, 1995.

137. Somoskovi A, Hotaling JE, Fitzgerald M et al. Lessons from a proficiency testing event for acid-fast microscopy. *Chest* 2001; **120**: 250-7.

138. Ba F, Rieder HL. A comparison of fluorescence microscopy with the Ziehl-Neelsen technique in the examination of sputum for acid-fast bacilli. *Int J Tuberc Lung Dis* 1999; **3**: 1101-5.

139. Peres MH. O Diagnóstico Microbiológico e os Novos Métodos de Diagnóstico Rápido. In: J. Pina, ed. *A Tuberculose na Viragem do Milénio*. Lisboa: Lidel, 2000.

140. Rickman TW, Moyer NP. Increased sensitivity of acid-fast smears. *J Clin Microbiol* 1980; **11**: 618-20.

141. Middlebrook G, Cohn ML. Bacteriology of tuberculosis: laboratory methods. *Am J Public Health Nations Health* 1958; **48**: 844-53.

142. Pfyffer GE, Welscher HM, Kissling P et al. Comparison of the Mycobacteria Growth Indicator Tube (MGIT) with radiometric and solid culture for recovery of acid-fast bacilli. *J Clin Microbiol* 1997; **35**: 364-8.

143. Somoskovi A, Kodmon C, Lantos A et al. Comparison of recoveries of mycobacterium tuberculosis using the automated BACTEC MGIT 960 system, the BACTEC 460 TB system, and Lowenstein-Jensen medium. *J Clin Microbiol* 2000; **38**: 2395-7.

144. Tortoli E, Cichero P, Piersimoni C et al. Use of BACTEC MGIT 960 for recovery of mycobacteria from clinical specimens: multicenter study. *J Clin Microbiol* 1999; **37**: 3578-82.

145. Canetti G, Froman S, Grosset J et al. Mycobacteria: Laboratory Methods for Testing Drug Sensitivity and Resistance. *Bull World Health Organ* 1963; **29**: 565-78.

146. NCCLS. Susceptibility Testing of Mycobacteria, Nocardiae, and Other Aerobic Actinomycetes; Approved Standard. NCCLS, 940 West Valley Road, Suite 1400, Wayne Pennsylvania 19087-1898 USA, 2003.

147. Canetti G, Fox W, Khomenko A et al. Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. *Bull World Health Organ* 1969; **41**: 21-43.

148. Bottger EC. The ins and outs of Mycobacterium tuberculosis drug susceptibility testing. *Clin Microbiol Infect* 2011; **17**: 1128-34.

149. Mitchison DA. Drug resistance in tuberculosis. *Eur Respir J* 2005; **25**: 376-9.

150. Zhang Y, Mitchison D. The curious characteristics of pyrazinamide: a review. *Int J Tuberc Lung Dis* 2003; **7**: 6-21.

151. Springer B, Lucke K, Calligaris-Maibach R et al. Quantitative drug susceptibility testing of Mycobacterium tuberculosis by use of MGIT 960 and EpiCenter instrumentation. *J Clin Microbiol* 2009; **47**: 1773-80.

152. Papaventsis D, Ioannidis P, Karabela S et al. Impact of the Gen-Probe Amplified MTD(R) Test on tuberculosis diagnosis in children. *Int J Tuberc Lung Dis* 2012; **16**: 384-90.

153. Fabre M, Vong R, Zrara A et al. [Performances of the Amplified Mycobacterium Tuberculosis Direct Test in non respiratory specimens (study about 1538 samples)]. *Pathol Biol (Paris)* 2011; **59**: 29-31.

154. Syre H, Myneedu VP, Arora VK et al. Direct detection of mycobacterial species in pulmonary specimens by two rapid amplification tests, the gen-probe amplified mycobacterium tuberculosis direct test and the genotype mycobacteria direct test. *J Clin Microbiol* 2009; **47**: 3635-9.

155. Neonakis IK, Gitti Z, Baritaki S et al. Evaluation of GenoType mycobacteria direct assay in comparison with Gen-Probe Mycobacterium tuberculosis amplified direct test and GenoType MTBDRplus for direct detection of Mycobacterium tuberculosis complex in clinical samples. *J Clin Microbiol* 2009; **47**: 2601-3.

156. Kalafati-Tzimaka E, Georgitzikis I, Delikatzi D et al. Use of the AMPLIFIED Mycobacterium tuberculosis direct test for the diagnosis of tuberculosis. *East Mediterr Health J* 2008; **14**: 1119-25.

157. el-Sayed Zaki M, Abou-el Hassan S. Clinical evaluation of Gen-Probe's amplified mycobacterium tuberculosis direct test for rapid diagnosis of Mycobacterium tuberculosis in Egyptian children at risk for infection. *Arch Pathol Lab Med* 2008; **132**: 244-7.

158. Guerra RL, Hooper NM, Baker JF et al. Use of the amplified mycobacterium tuberculosis direct test in a public health laboratory: test performance and impact on clinical care. *Chest* 2007; **132**: 946-51.

159. Pollock N, Westerling J, Sloutsky A. Specimen dilution increases the diagnostic utility of the gen-probe mycobacterium tuberculosis direct test. *Am J Clin Pathol* 2006; **126**: 142-7.

160. Kerleguer A, Fabre M, Bernatas JJ et al. Clinical evaluation of the gen-probe amplified mycobacterium tuberculosis direct test for rapid diagnosis of tuberculosis lymphadenitis. *J Clin Microbiol* 2004; **42**: 5921-2.

161. Lemaitre N, Armand S, Vachee A et al. Comparison of the real-time PCR method and the Gen-Probe amplified Mycobacterium tuberculosis direct test for detection of Mycobacterium tuberculosis in pulmonary and nonpulmonary specimens. *J Clin Microbiol* 2004; **42**: 4307-9.

162. Pounder JI, Aldous WK, Woods GL. Comparison of real-time polymerase chain reaction using the Smart Cycler and the Gen-Probe amplified Mycobacterium tuberculosis direct test for detection of M. tuberculosis complex in clinical specimens. *Diagn Microbiol Infect Dis* 2006; **54**: 217-22.

163. Marttila HJ, Soini H, Vyshnevskaya E et al. Line probe assay in the rapid detection of rifampin-resistant Mycobacterium tuberculosis directly from clinical specimens. *Scand J Infect Dis* 1999; **31**: 269-73.

164. Sam IC, Drobniewski F, More P et al. Mycobacterium tuberculosis and rifampin resistance, United Kingdom. *Emerg Infect Dis* 2006; **12**: 752-9.

165. Tortoli E, Marcelli F. Use of the INNO LiPA Rif.TB for detection of Mycobacterium tuberculosis DNA directly in clinical specimens and for simultaneous determination of rifampin susceptibility. *Eur J Clin Microbiol Infect Dis* 2007; **26**: 51-5.

166. Traore H, van Deun A, Shamputa IC et al. Direct detection of Mycobacterium tuberculosis complex DNA and rifampin resistance in clinical specimens from tuberculosis patients by line probe assay. *J Clin Microbiol* 2006; **44**: 4384-8.

167. Viveiros M, Leandro C, Rodrigues L et al. Direct application of the INNO-LiPA Rif.TB line-probe assay for rapid identification of Mycobacterium tuberculosis complex strains and detection of rifampin resistance in 360 smear-positive respiratory specimens from an area of high incidence of multidrug-resistant tuberculosis. *J Clin Microbiol* 2005; **43**: 4880-4.

168. Bang D, Bengard Andersen A, Thomsen VO. Rapid genotypic detection of rifampin- and isoniazid-resistant Mycobacterium tuberculosis directly in clinical specimens. *J Clin Microbiol* 2006; **44**: 2605-8.

169. Friedrich SO, Venter A, Kayigire XA et al. Suitability of Xpert MTB/RIF and genotype MTBDRplus for patient selection for a tuberculosis clinical trial. *J Clin Microbiol* 2011; **49**: 2827-31.

170. Hillemann D, Rusch-Gerdes S, Richter E. Application of the Genotype MTBDR assay directly on sputum specimens. *Int J Tuberc Lung Dis* 2006; **10**: 1057-9.

171. Somoskovi A, Dormandy J, Mitsani D et al. Use of smear-positive samples to assess the PCR-based genotype MTBDR assay for rapid, direct detection of the Mycobacterium tuberculosis complex as well as its resistance to isoniazid and rifampin. *J Clin Microbiol* 2006; **44**: 4459-63.

172. Lawn SD, Nicol MP. Xpert(R) MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiol* 2011; **6**: 1067-82.

173. Dorman SE, Chihota VN, Lewis JJ et al. Performance Characteristics of the Cepheid Xpert MTB/RIF Test in a Tuberculosis Prevalence Survey. *PLoS ONE* 2012; **7**: e43307.

174. Balcells ME, Garcia P, Chanqueo L et al. Rapid molecular detection of pulmonary tuberculosis in HIV-infected patients in Santiago, Chile. *Int J Tuberc Lung Dis* 2012.

175. Kirwan DE, Cardenas MK, Gilman RH. Rapid Implementation of New TB Diagnostic Tests: Is It Too Soon for a Global Roll-Out of Xpert MTB/RIF? *Am J Trop Med Hyg* 2012; **87**: 197-201.

176. Williamson DA, Basu I, Bower J et al. An evaluation of the Xpert MTB/RIF assay and detection of false-positive rifampicin resistance in Mycobacterium tuberculosis. *Diagn Microbiol Infect Dis* 2012.

177. Peter JG, Theron G, Muchinga TE et al. The Diagnostic Accuracy of Urine-Based Xpert MTB/RIF in HIV-Infected Hospitalized Patients Who Are Smear-Negative or Sputum Scarce. *PLoS ONE* 2012; **7**: e39966.

178. O'Grady J, Bates M, Chilukutu L et al. Evaluation of the Xpert(R) MTB/RIF assay at a tertiary referral hospital in a high TB/HIV endemic setting. *Clin Infect Dis* 2012.

179. Rachow A, Clowes P, Saathoff E et al. Increased and expedited case detection by Xpert MTB/RIF assay in childhood tuberculosis: a prospective cohort study. *Clin Infect Dis* 2012; **54**: 1388-96.

180. Chang K, Lu W, Wang J et al. Rapid and effective diagnosis of tuberculosis and rifampicin resistance with Xpert MTB/RIF assay: a meta-analysis. *J Infect* 2012; **64**: 580-8.

181. Safianowska A, Walkiewicz R, Nejman-Gryz P et al. [Two selected commercially based nucleic acid amplification tests for the diagnosis of tuberculosis]. *Pneumonol Alergol Pol* 2012; **80**: 6-12.

182. Tortoli E, Russo C, Piersimoni C et al. Clinical validation of Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis. *Eur Respir J* 2012; **40**: 442-7.

183. Moure R, Martin R, Alcaide F. Effectiveness of an integrated real-time PCR method for detection of the Mycobacterium tuberculosis complex in smear-negative extrapulmonary samples in an area of low tuberculosis prevalence. *J Clin Microbiol* 2012; **50**: 513-5.

184. Friedrich SO, von Groote-Bidlingmaier F, Diacon AH. Xpert MTB/RIF assay for diagnosis of pleural tuberculosis. *J Clin Microbiol* 2011; **49**: 4341-2.

185. Teo J, Jureen R, Chiang D et al. Comparison of two nucleic acid amplification assays, the Xpert MTB/RIF assay and the amplified Mycobacterium Tuberculosis Direct assay, for detection of Mycobacterium tuberculosis in respiratory and nonrespiratory specimens. *J Clin Microbiol* 2011; **49**: 3659-62.

186. Miller MB, Popowitch EB, Backlund MG et al. Performance of Xpert MTB/RIF RUO assay and IS6110 real-time PCR for Mycobacterium tuberculosis detection in clinical samples. *J Clin Microbiol* 2011; **49**: 3458-62.

187. Scott LE, McCarthy K, Gous N et al. Comparison of Xpert MTB/RIF with other nucleic acid technologies for diagnosing pulmonary tuberculosis in a high HIV prevalence setting: a prospective study. *PLoS Med* 2011; **8**: e1001061.

188. Rachow A, Zumla A, Heinrich N et al. Rapid and accurate detection of Mycobacterium tuberculosis in sputum samples by Cepheid Xpert MTB/RIF assay--a clinical validation study. *PLoS ONE* 2011; **6**: e20458.

189. Skenders GK, Holtz TH, Riekstina V et al. Implementation of the INNO-LiPA Rif. TB(R) line-probe assay in rapid detection of multidrug-resistant tuberculosis in Latvia. *Int J Tuberc Lung Dis* 2011; **15**: 1546-52, i.

190. Seibert F. The isolation and properties of the purified protein derivative of tuberculin. *Am Rev Tuberc* 1934; **30**: 713-20.

191. Santin M, Munoz L, Rigau D. Interferon-gamma release assays for the diagnosis of tuberculosis and tuberculosis infection in HIV-infected adults: a systematic review and meta-analysis. *PLoS ONE* 2012; **7**: e32482.

192. Cattamanchi A, Smith R, Steingart KR et al. Interferon-gamma release assays for the diagnosis of latent tuberculosis infection in HIV-infected individuals: a systematic review and meta-analysis. *J Acquir Immune Defic Syndr* 2011; **56**: 230-8.

193. Chen J, Zhang R, Wang J et al. Interferon-gamma release assays for the diagnosis of active tuberculosis in HIV-infected patients: a systematic review and meta-analysis. *PLoS ONE* 2011; **6**: e26827.

194. Metcalfe JZ, Everett CK, Steingart KR et al. Interferon-gamma release assays for active pulmonary tuberculosis diagnosis in adults in low- and middle-income countries: systematic review and meta-analysis. *J Infect Dis* 2011; **204 Suppl 4**: S1120-9.

195. Banfield S, Pascoe E, Thambiran A et al. Factors associated with the performance of a blood-based interferon-gamma release assay in diagnosing tuberculosis. *PLoS ONE* 2012; **7**: e38556.

196. Hang NT, Lien LT, Kobayashi N et al. Analysis of factors lowering sensitivity of interferon-gamma release assay for tuberculosis. *PLoS ONE* 2011; **6**: e23806.

197. Aichelburg MC, Rieger A, Breitenecker F et al. Detection and prediction of active tuberculosis disease by a whole-blood interferon-gamma release assay in HIV-1-infected individuals. *Clin Infect Dis* 2009; **48**: 954-62.

198. Santin M, Casas S, Saumoy M et al. Detection of latent tuberculosis by the tuberculin skin test and a whole-blood interferon-gamma release assay, and the development of active tuberculosis in HIV-seropositive persons. *Diagn Microbiol Infect Dis* 2011; **69**: 59-65.

199. Mitchison DA. Basic mechanisms of chemotherapy. *Chest* 1979; **76**: 771-81.

200. Jawahar MS. Current trends in chemotherapy of tuberculosis. *Indian J Med Res* 2004; **120**: 398-417.

201. Kadhiravan T, Sharma SK. Medical management of genitourinary tuberculosis. *Indian J Urol* 2008; **24**: 362-8.

202. World Health Organization. Tratamento da tuberculose: linhas orientadoras para programas nacionais. DGS, 2004. Original Title: Treatment of tuberculosis: guidelines for national programmes, 3rd edition. Lisbon. 2004.

203. World Health Organization. Treatment of tuberculosis : guidelines - 4th ed. Geneva. World Health Organization. 2010. WHO Publication No. WHO/HTM/TB/2009.420.

204. Heym B, Zhang Y, Poulet S et al. Characterization of the katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of Mycobacterium tuberculosis. *J Bacteriol* 1993; **175**: 4255-9.

205. Rozwarski DA, Grant GA, Barton DH et al. Modification of the NADH of the isoniazid target (InhA) from Mycobacterium tuberculosis. *Science (New York, NY* 1998; **279**: 98-102.

206. Vilcheze C, Wang F, Arai M et al. Transfer of a point mutation in Mycobacterium tuberculosis inhA resolves the target of isoniazid. *Nat Med* 2006; **12**: 1027-9.

207. Vilcheze C, Weisbrod TR, Chen B et al. Altered NADH/NAD+ ratio mediates coresistance to isoniazid and ethionamide in mycobacteria. *Antimicrob Agents Chemother* 2005; **49**: 708-20.

208. Bardou F, Raynaud C, Ramos C et al. Mechanism of isoniazid uptake in Mycobacterium tuberculosis. *Microbiology* 1998; **144 (Pt 9)**: 2539-44.

209. Rastogi N, Labrousse V, Goh KS. In vitro activities of fourteen antimicrobial agents against drug susceptible and resistant clinical isolates of Mycobacterium tuberculosis and comparative intracellular activities against the virulent H37Rv strain in human macrophages. *Curr Microbiol* 1996; **33**: 167-75.

210. Ando H, Kondo Y, Suetake T et al. Identification of katG mutations associated with high-level isoniazid resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2010; **54**: 1793-9.

211. Hausner TP, Geigenmuller U, Nierhaus KH. The allosteric three-site model for the ribosomal elongation cycle. New insights into the inhibition mechanisms of aminoglycosides, thiostrepton, and viomycin. *J Biol Chem* 1988; **263**: 13103-11.

212. Zhang Y, Garbe T, Young D. Transformation with katG restores isoniazid-sensitivity in Mycobacterium tuberculosis isolates resistant to a range of drug concentrations. *Mol Microbiol* 1993; **8**: 521-4.

213. Zhang Y, Heym B, Allen B et al. The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. *Nature* 1992; **358**: 591-3.

214. Feuerriegel S, Oberhauser B, George AG et al. Sequence analysis for detection of first-line drug resistance in Mycobacterium tuberculosis strains from a high-incidence setting. *BMC Microbiol* 2012; **12**: 90.

215. Yuan X, Zhang T, Kawakami K et al. Molecular characterization of multidrug- and extensively drug-resistant Mycobacterium tuberculosis strains in Jiangxi, China. *J Clin Microbiol* 2012; **50**: 2404-13.

216. Afanas'ev MV, Ikryannikova LN, Il'ina EN et al. Molecular characteristics of rifampicin- and isoniazid-resistant Mycobacterium tuberculosis isolates from the Russian Federation. *J Antimicrob Chemother* 2007; **59**: 1057-64.

217. Alves SL, Metzker FS, Araujo-Filho JA et al. Clinical data and molecular analysis of Mycobacterium tuberculosis isolates from drug-resistant tuberculosis patients in Goias, Brazil. *Mem Inst Oswaldo Cruz* 2011; **106**: 655-61.

218. Bakonyte D, Baranauskaite A, Cicenaite J et al. Molecular characterization of isoniazid-resistant Mycobacterium tuberculosis clinical isolates in Lithuania. *Antimicrob Agents Chemother* 2003; **47**: 2009-11.

219. Campbell PJ, Morlock GP, Sikes RD et al. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2011; **55**: 2032-41.

220. Chaoui I, Sabouni R, Kourout M et al. Analysis of isoniazid, streptomycin and ethambutol resistance in Mycobacterium tuberculosis isolates from Morocco. *J Infect Dev Ctries* 2009; **3**: 278-84.

221. Cho EH, Bae HK, Kang SK et al. Detection of isoniazid and rifampicin resistance by sequencing of katG, inhA, and rpoB genes in Korea. *Korean J Lab Med* 2009; **29**: 455-60.

222. Dalla Costa ER, Ribeiro MO, Silva MS et al. Correlations of mutations in katG, oxyR-ahpC and inhA genes and in vitro susceptibility in Mycobacterium tuberculosis clinical strains segregated by spoligotype families from tuberculosis prevalent countries in South America. *BMC Microbiol* 2009; **9**: 39.

223. Homolka S, Meyer CG, Hillemann D et al. Unequal distribution of resistance-conferring mutations among Mycobacterium tuberculosis and Mycobacterium africanum strains from Ghana. *Int J Med Microbiol* 2010; **300**: 489-95.

224. Minh NN, Van Bac N, Son NT et al. Molecular characteristics of rifampin- and isoniazid-resistant mycobacterium tuberculosis strains isolated in Vietnam. *J Clin Microbiol* 2012; **50**: 598-601.

225. Muller B, Streicher EM, Hoek KG et al. inhA promoter mutations: a gateway to extensively drug-resistant tuberculosis in South Africa? *Int J Tuberc Lung Dis* 2011; **15**: 344-51.

226. Nikolayevsky V, Brown T, Balabanova Y et al. Detection of mutations associated with isoniazid and rifampin resistance in Mycobacterium tuberculosis isolates from Samara Region, Russian Federation. *J Clin Microbiol* 2004; **42**: 4498-502.

227. Rindi L, Bianchi L, Tortoli E et al. Mutations responsible for Mycobacterium tuberculosis isoniazid resistance in Italy. *Int J Tuberc Lung Dis* 2005; **9**: 94-7.

228. Sajduda A, Brzostek A, Poplawska M et al. Molecular characterization of rifampin- and isoniazid-resistant Mycobacterium tuberculosis strains isolated in Poland. *J Clin Microbiol* 2004; **42**: 2425-31.

229. Silva MS, Senna SG, Ribeiro MO et al. Mutations in katG, inhA, and ahpC genes of Brazilian isoniazid-resistant isolates of Mycobacterium tuberculosis. *J Clin Microbiol* 2003; **41**: 4471-4.

230. Valvatne H, Syre H, Kross M et al. Isoniazid and rifampicin resistance-associated mutations in Mycobacterium tuberculosis isolates from Yangon, Myanmar: implications for rapid molecular testing. *J Antimicrob Chemother* 2009; **64**: 694-701.

231. Yao C, Zhu T, Li Y et al. Detection of rpoB, katG and inhA gene mutations in Mycobacterium tuberculosis clinical isolates from Chongqing as determined by microarray. *Clin Microbiol Infect* 2010; **16**: 1639-43.

232. Zhang M, Yue J, Yang YP et al. Detection of mutations associated with isoniazid resistance in Mycobacterium tuberculosis isolates from China. *J Clin Microbiol* 2005; **43**: 5477-82.

233. Chan RC, Hui M, Chan EW et al. Genetic and phenotypic characterization of drug-resistant Mycobacterium tuberculosis isolates in Hong Kong. *J Antimicrob Chemother* 2007; **59**: 866-73.

234. Ali A, Hasan R, Jabeen K et al. Characterization of mutations conferring extensive drug resistance to Mycobacterium tuberculosis isolates in Pakistan. *Antimicrob Agents Chemother* 2011; **55**: 5654-9.

235. Saint-Joanis B, Souchon H, Wilming M et al. Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from Mycobacterium tuberculosis. *Biochem J* 1999; **338 (Pt 3)**: 753-60.

236. Wengenack NL, Todorovic S, Yu L et al. Evidence for differential binding of isoniazid by Mycobacterium tuberculosis KatG and the isoniazid-resistant mutant KatG(S315T). *Biochemistry* 1998; **37**: 15825-34.

237. Yu S, Girotto S, Lee C et al. Reduced affinity for Isoniazid in the S315T mutant of Mycobacterium tuberculosis KatG is a key factor in antibiotic resistance. *J Biol Chem* 2003; **278**: 14769-75.

238. Larsen MH, Vilcheze C, Kremer L et al. Overexpression of inhA, but not kasA, confers resistance to isoniazid and ethionamide in Mycobacterium smegmatis, M. bovis BCG and M. tuberculosis. *Mol Microbiol* 2002; **46**: 453-66.

239. Cardoso RF, Cardoso MA, Leite CQ et al. Characterization of ndh gene of isoniazid resistant and susceptible Mycobacterium tuberculosis isolates from Brazil. *Mem Inst Oswaldo Cruz* 2007; **102**: 59-61.

240. Ho YM, Sun YJ, Wong SY et al. Contribution of dfrA and inhA mutations to the detection of isoniazid-resistant Mycobacterium tuberculosis isolates. *Antimicrob Agents Chemother* 2009; **53**: 4010-2.

241. Machado D, Couto I, Perdigao J et al. Contribution of efflux to the emergence of isoniazid and multidrug resistance in Mycobacterium tuberculosis. *PLoS ONE* 2012; **7**: e34538.

242. Coll P. Fármacos con actividad frente a Mycobacterium tuberculosis. *Enferm Infecc Microbiol Clin* 2003; **21**: 299–308.

243. Wehrli W, Knusel F, Schmid K et al. Interaction of rifamycin with bacterial RNA polymerase. *Proc Natl Acad Sci U S A* 1968; **61**: 667-73.

244. McClure WR, Cech CL. On the mechanism of rifampicin inhibition of RNA synthesis. *J Biol Chem* 1978; **253**: 8949-56.

245. Campbell EA, Korzheva N, Mustaev A et al. Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 2001; **104**: 901-12.

246. Handbook of anti-tuberculosis agents. Introduction. *Tuberculosis (Edinburgh, Scotland)* 2008; **88**: 85-6.

247. Stottmeier KD, Kubica GP, Woodley CL. Antimycobacterial activity of rifampin under in vitro and simulated in vivo conditions. *Appl Microbiol* 1969; **17**: 861-5.

248. Gillespie SH. Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. *Antimicrob Agents Chemother* 2002; **46**: 267-74.

249. Heep M, Brandstatter B, Rieger U et al. Frequency of rpoB mutations inside and outside the cluster I region in rifampin-resistant clinical Mycobacterium tuberculosis isolates. *J Clin Microbiol* 2001; **39**: 107-10.

250. Kapur V, Li LL, Iordanescu S et al. Characterization by automated DNA sequencing of mutations in the gene (rpoB) encoding the RNA polymerase beta subunit in rifampin-resistant Mycobacterium tuberculosis strains from New York City and Texas. *J Clin Microbiol* 1994; **32**: 1095-8.

251. Miller LP, Crawford JT, Shinnick TM. The rpoB gene of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 1994; **38**: 805-11.

252. Herrera L, Jimenez S, Valverde A et al. Molecular analysis of rifampicin-resistant Mycobacterium tuberculosis isolated in Spain (1996-2001). Description of new mutations in the rpoB gene and review of the literature. *Int J Antimicrob Agents* 2003; **21**: 403-8.

253. Suresh N, Singh UB, Arora J et al. rpoB gene sequencing and spoligotyping of multidrug-resistant Mycobacterium tuberculosis isolates from India. *Infect Genet Evol* 2006; **6**: 474-83.

254. Ahmad S, Mokaddas E. The occurrence of rare rpoB mutations in rifampicin-resistant clinical Mycobacterium tuberculosis isolates from Kuwait. *Int J Antimicrob Agents* 2005; **26**: 205-12.

255. Aristimuno L, Armengol R, Cebollada A et al. Molecular characterisation of Mycobacterium tuberculosis isolates in the First National Survey of Anti-tuberculosis Drug Resistance from Venezuela. *BMC Microbiol* 2006; **6**: 90.

256. Bakonyte D, Baranauskaite A, Cicenaite J et al. Mutations in the rpoB gene of rifampicin-resistant Mycobacterium tuberculosis clinical isolates from Lithuania. *Int J Tuberc Lung Dis* 2005; **9**: 936-8.

257. Lee AS, Lim IH, Tang LL et al. High frequency of mutations in the rpoB gene in rifampin-resistant clinical isolates of Mycobacterium tuberculosis from Singapore. *J Clin Microbiol* 2005; **43**: 2026-7.

258. O'Sullivan DM, McHugh TD, Gillespie SH. Analysis of rpoB and pncA mutations in the published literature: an insight into the role of oxidative stress in Mycobacterium tuberculosis evolution? *J Antimicrob Chemother* 2005; **55**: 674-9.

259. Rahim Z, Nakajima C, Raqib R et al. Molecular mechanism of rifampicin and isoniazid resistance in Mycobacterium tuberculosis from Bangladesh. *Tuberculosis (Edinburgh, Scotland)* 2012.

260. Shemyakin IG, Stepanshina VN, Ivanov IY et al. Characterization of drug-resistant isolates of Mycobacterium tuberculosis derived from Russian inmates. *Int J Tuberc Lung Dis* 2004; **8**: 1194-203.

261. Siu GK, Zhang Y, Lau TC et al. Mutations outside the rifampicin resistance-determining region associated with rifampicin resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother* 2011; **66**: 730-3.

262. Gagneux S, Long CD, Small PM et al. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. *Science (New York, NY* 2006; **312**: 1944-6.

263. Mariam DH, Mengistu Y, Hoffner SE et al. Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2004; **48**: 1289-94.

264. van Ingen J, Aarnoutse R, de Vries G et al. Low-level rifampicin-resistant Mycobacterium tuberculosis strains raise a new therapeutic challenge. *Int J Tuberc Lung Dis* 2011; **15**: 990-2.

265. Van Deun A, Barrera L, Bastian I et al. Mycobacterium tuberculosis strains with highly discordant rifampin susceptibility test results. *J Clin Microbiol* 2009; **47**: 3501-6.

266. Zaczek A, Brzostek A, Augustynowicz-Kopec E et al. Genetic evaluation of relationship between mutations in rpoB and resistance of Mycobacterium tuberculosis to rifampin. *BMC Microbiol* 2009; **9**: 10.

267. Quan S, Venter H, Dabbs ER. Ribosylative inactivation of rifampin by Mycobacterium smegmatis is a principal contributor to its low susceptibility to this antibiotic. *Antimicrob Agents Chemother* 1997; **41**: 2456-60.

268. Dabbs ER, Yazawa K, Mikami Y et al. Ribosylation by mycobacterial strains as a new mechanism of rifampin inactivation. *Antimicrob Agents Chemother* 1995; **39**: 1007-9.

269. Takayama K, Kilburn JO. Inhibition of synthesis of arabinogalactan by ethambutol in Mycobacterium smegmatis. *Antimicrob Agents Chemother* 1989; **33**: 1493-9.

270. Mikusova K, Slayden RA, Besra GS et al. Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. *Antimicrob Agents Chemother* 1995; **39**: 2484-9.

271. Shi L, Berg S, Lee A et al. The carboxy terminus of EmbC from Mycobacterium smegmatis mediates chain length extension of the arabinan in lipoarabinomannan. *J Biol Chem* 2006; **281**: 19512-26.

272. Zhang N, Torrelles JB, McNeil MR et al. The Emb proteins of mycobacteria direct arabinosylation of lipoarabinomannan and arabinogalactan via an N-terminal recognition region and a C-terminal synthetic region. *Mol Microbiol* 2003; **50**: 69-76.

273. Goude R, Amin AG, Chatterjee D et al. The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2009; **53**: 4138-46.

274. Beggs WH, Auran NE. Uptake and binding of 14C-ethambutol by tubercle bacilli and the relation of binding to growth inhibition. *Antimicrob Agents Chemother* 1972; **2**: 390-4.

275. Forbes M, Kuck NA, Peets EA. Mode of action of ethambutol. *J Bacteriol* 1962; **84**: 1099-103.

276. Bakker-Woudenberg IA, van Vianen W, van Soolingen D et al. Antimycobacterial agents differ with respect to their bacteriostatic versus bactericidal activities in relation to time of exposure, mycobacterial growth phase, and their use in combination. *Antimicrob Agents Chemother* 2005; **49**: 2387-98.

277. Alcaide F, Pfyffer GE, Telenti A. Role of embB in natural and acquired resistance to ethambutol in mycobacteria. *Antimicrob Agents Chemother* 1997; **41**: 2270-3.

278. Ramaswamy SV, Amin AG, Goksel S et al. Molecular genetic analysis of nucleotide polymorphisms associated with ethambutol resistance in human isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2000; **44**: 326-36.

279. Telenti A, Philipp WJ, Sreevatsan S et al. The emb operon, a gene cluster of Mycobacterium tuberculosis involved in resistance to ethambutol. *Nat Med* 1997; **3**: 567-70.

280. Lety MA, Nair S, Berche P et al. A single point mutation in the embB gene is responsible for resistance to ethambutol in Mycobacterium smegmatis. *Antimicrob Agents Chemother* 1997; **41**: 2629-33.

281. Jadaun GP, Das R, Upadhyay P et al. Role of embCAB gene mutations in ethambutol resistance in Mycobacterium tuberculosis isolates from India. *Int J Antimicrob Agents* 2009; **33**: 483-6.

282. Lee HY, Myoung HJ, Bang HE et al. Mutations in the embB locus among Korean clinical isolates of Mycobacterium tuberculosis resistant to ethambutol. *Yonsei Med J* 2002; **43**: 59-64.

283. Mokrousov I, Otten T, Vyshnevskiy B et al. Detection of embB306 mutations in ethambutol-susceptible clinical isolates of Mycobacterium tuberculosis from Northwestern Russia: implications for genotypic resistance testing. *J Clin Microbiol* 2002; **40**: 3810-3.

284. Plinke C, Rusch-Gerdes S, Niemann S. Significance of mutations in embB codon 306 for prediction of ethambutol resistance in clinical Mycobacterium tuberculosis isolates. *Antimicrob Agents Chemother* 2006; **50**: 1900-2.

285. Sreevatsan S, Stockbauer KE, Pan X et al. Ethambutol resistance in Mycobacterium tuberculosis: critical role of embB mutations. *Antimicrob Agents Chemother* 1997; **41**: 1677-81.

286. Srivastava S, Ayyagari A, Dhole TN et al. emb nucleotide polymorphisms and the role of embB306 mutations in Mycobacterium tuberculosis resistance to ethambutol. *Int J Med Microbiol* 2009; **299**: 269-80.

287. Tracevska T, Jansone I, Nodieva A et al. Characterisation of rpsL, rrs and embB mutations associated with streptomycin and ethambutol resistance in Mycobacterium tuberculosis. *Res Microbiol* 2004; **155**: 830-4.

288. Ahmad S, Jaber AA, Mokaddas E. Frequency of embB codon 306 mutations in ethambutol-susceptible and -resistant clinical Mycobacterium tuberculosis isolates in Kuwait. *Tuberculosis (Edinburgh, Scotland)* 2007; **87**: 123-9.

289. Parsons LM, Salfinger M, Clobridge A et al. Phenotypic and molecular characterization of Mycobacterium tuberculosis isolates resistant to both isoniazid and ethambutol. *Antimicrob Agents Chemother* 2005; **49**: 2218-25.

290. Srivastava S, Garg A, Ayyagari A et al. Nucleotide polymorphism associated with ethambutol resistance in clinical isolates of Mycobacterium tuberculosis. *Curr Microbiol* 2006; **53**: 401-5.

291. Park YK, Ryoo SW, Lee SH et al. Correlation of the phenotypic ethambutol susceptibility of Mycobacterium tuberculosis with embB gene mutations in Korea. *J Med Microbiol* 2012; **61**: 529-34.

292. Johnson R, Jordaan AM, Pretorius L et al. Ethambutol resistance testing by mutation detection. *Int J Tuberc Lung Dis* 2006; **10**: 68-73.

293. Shen X, Shen GM, Wu J et al. Association between embB codon 306 mutations and drug resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2007; **51**: 2618-20.

294. Plinke C, Cox HS, Kalon S et al. Tuberculosis ethambutol resistance: concordance between phenotypic and genotypic test results. *Tuberculosis (Edinburgh, Scotland)* 2009; **89**: 448-52.

295. Plinke C, Walter K, Aly S et al. Mycobacterium tuberculosis embB codon 306 mutations confer moderately increased resistance to ethambutol in vitro and in vivo. *Antimicrob Agents Chemother* 2011; **55**: 2891-6.

296. Safi H, Fleischmann RD, Peterson SN et al. Allelic exchange and mutant selection demonstrate that common clinical embCAB gene mutations only modestly increase resistance to ethambutol in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2010; **54**: 103-8.

297. Safi H, Sayers B, Hazbon MH et al. Transfer of embB codon 306 mutations into clinical Mycobacterium tuberculosis strains alters susceptibility to ethambutol, isoniazid, and rifampin. *Antimicrob Agents Chemother* 2008; **52**: 2027-34.

298. Colangeli R, Helb D, Sridharan S et al. The Mycobacterium tuberculosis iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Mol Microbiol* 2005; **55**: 1829-40.

299. Jaber AA, Ahmad S, Mokaddas E. Minor contribution of mutations at iniA codon 501 and embC-embA intergenic region in ethambutol-resistant clinical Mycobacterium tuberculosis isolates in Kuwait. *Ann Clin Microbiol Antimicrob* 2009; **8**: 2.

300. Waksman SA, Reilly HC, Johnstone DB. Isolation of Streptomycin-producing Strains of Streptomyces griseus. *J Bacteriol* 1946; **52**: 393-7.

301. Smith DG, Waksman SA. Tuberculostatic and Tuberculocidal Properties of Streptomycin. *J Bacteriol* 1947; **54**: 253-61.

302. Carter AP, Clemons WM, Brodersen DE et al. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 2000; **407**: 340-8.

303. Ruusala T, Kurland CG. Streptomycin preferentially perturbs ribosomal proofreading. *Mol Gen Genet* 1984; **198**: 100-4.

304. Davis BD, Chen LL, Tai PC. Misread protein creates membrane channels: an essential step in the bactericidal action of aminoglycosides. *Proc Natl Acad Sci U S A* 1986; **83**: 6164-8.

305. Bryan LE, Kwan S. Roles of ribosomal binding, membrane potential, and electron transport in bacterial uptake of streptomycin and gentamicin. *Antimicrob Agents Chemother* 1983; **23**: 835-45.

306. Busse HJ, Wostmann C, Bakker EP. The bactericidal action of streptomycin: membrane permeabilization caused by the insertion of mistranslated proteins into the cytoplasmic membrane of Escherichia coli and subsequent caging of the antibiotic inside the cells due to degradation of these proteins. *J Gen Microbiol* 1992; **138**: 551-61.

307. Hewitt J, Kogut M. An investigation of mistranslation in vivo induced by streptomycin by an examination of the susceptibility of abnormal proteins to degradation. *Eur J Biochem* 1977; **74**: 285-92.

308. Karimi R, Ehrenberg M. Dissociation rate of cognate peptidyl-tRNA from the A-site of hyper-accurate and error-prone ribosomes. *Eur J Biochem* 1994; **226**: 355-60.

309. Ruiz P, Rodriguez-Cano F, Zerolo FJ et al. [Streptomycin as second-line chemotherapy for tuberculosis]. *Rev Esp Quimioter* 2003; **16**: 188-94.

310. Peloquin CA, Berning SE, Nitta AT et al. Aminoglycoside toxicity: daily versus thrice-weekly dosing for treatment of mycobacterial diseases. *Clin Infect Dis* 2004; **38**: 1538-44.

311. Meier A, Kirschner P, Bange FC et al. Genetic alterations in streptomycin-resistant Mycobacterium tuberculosis: mapping of mutations conferring resistance. *Antimicrob Agents Chemother* 1994; **38**: 228-33.

312. Cooksey RC, Morlock GP, McQueen A et al. Characterization of streptomycin resistance mechanisms among Mycobacterium tuberculosis isolates from patients in New York City. *Antimicrob Agents Chemother* 1996; **40**: 1186-8.

313. Meier A, Sander P, Schaper KJ et al. Correlation of molecular resistance mechanisms and phenotypic resistance levels in streptomycin-resistant Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 1996; **40**: 2452-4.

314. Springer B, Kidan YG, Prammananan T et al. Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrob Agents Chemother* 2001; **45**: 2877-84.

315. Tudo G, Rey E, Borrell S et al. Characterization of mutations in streptomycin-resistant Mycobacterium tuberculosis clinical isolates in the area of Barcelona. *J Antimicrob Chemother* 2010; **65**: 2341-6.

316. Brzostek A, Sajduda A, Sliwinski T et al. Molecular characterisation of streptomycin-resistant Mycobacterium tuberculosis strains isolated in Poland. *Int J Tuberc Lung Dis* 2004; **8**: 1032-5.

317. Cuevas-Cordoba B, Cuellar-Sanchez A, Pasissi-Crivelli A et al. rrs and rpsL mutations in streptomycin-resistant isolates of Mycobacterium tuberculosis from Mexico. *J Microbiol Immunol Infect* 2012.

318. Feuerriegel S, Oberhauser B, George AG et al. Sequence analysis for detection of first-line drug resistance in Mycobacterium tuberculosis strains from a high-incidence setting. *BMC Microbiol* 2012; **12**: 90.

319. Lipin MY, Stepanshina VN, Shemyakin IG et al. Association of specific mutations in katG, rpoB, rpsL and rrs genes with spoligotypes of multidrug-resistant Mycobacterium tuberculosis isolates in Russia. *Clin Microbiol Infect* 2007; **13**: 620-6.

320. Ramaswamy SV, Dou SJ, Rendon A et al. Genotypic analysis of multidrug-resistant Mycobacterium tuberculosis isolates from Monterrey, Mexico. *J Med Microbiol* 2004; **53**: 107-13.

321. Sekiguchi J, Miyoshi-Akiyama T, Augustynowicz-Kopec E et al. Detection of multidrug resistance in Mycobacterium tuberculosis. *J Clin Microbiol* 2007; **45**: 179-92.

322. Shi R, Zhang J, Li C et al. Detection of streptomycin resistance in Mycobacterium tuberculosis clinical isolates from China as determined by denaturing HPLC analysis and DNA sequencing. *Microbes Infect* 2007; **9**: 1538-44.

323. Sreevatsan S, Pan X, Stockbauer KE et al. Characterization of rpsL and rrs mutations in streptomycin-resistant Mycobacterium tuberculosis isolates from diverse geographic localities. *Antimicrob Agents Chemother* 1996; **40**: 1024-6.

324. Sun YJ, Luo JT, Wong SY et al. Analysis of rpsL and rrs mutations in Beijing and non-Beijing streptomycin-resistant Mycobacterium tuberculosis isolates from Singapore. *Clin Microbiol Infect* 2010; **16**: 287-9.

325. Okamoto S, Tamaru A, Nakajima C et al. Loss of a conserved 7-methylguanosine modification in 16S rRNA confers low-level streptomycin resistance in bacteria. *Mol Microbiol* 2007; **63**: 1096-106.

326. Spies FS, da Silva PE, Ribeiro MO et al. Identification of mutations related to streptomycin resistance in clinical isolates of Mycobacterium tuberculosis and possible involvement of efflux mechanism. *Antimicrob Agents Chemother* 2008; **52**: 2947-9.

327. Spies FS, Ribeiro AW, Ramos DF et al. Streptomycin resistance and lineage-specific polymorphisms in Mycobacterium tuberculosis gidB gene. *J Clin Microbiol* 2011; **49**: 2625-30.

328. Ballif M, Harino P, Ley S et al. Drug resistance-conferring mutations in Mycobacterium tuberculosis from Madang, Papua New Guinea. *BMC Microbiol* 2012; **12**: 191.

329. Zhang Y, Scorpio A, Nikaido H et al. Role of acid pH and deficient efflux of pyrazinoic acid in unique susceptibility of Mycobacterium tuberculosis to pyrazinamide. *J Bacteriol* 1999; **181**: 2044-9.

330. Zhang Y, Permar S, Sun Z. Conditions that may affect the results of susceptibility testing of Mycobacterium tuberculosis to pyrazinamide. *J Med Microbiol* 2002; **51**: 42-9.

331. Wade MM, Zhang Y. Anaerobic incubation conditions enhance pyrazinamide activity against Mycobacterium tuberculosis. *J Med Microbiol* 2004; **53**: 769-73.

332. Scorpio A, Lindholm-Levy P, Heifets L et al. Characterization of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 1997; **41**: 540-3.

333. Zimhony O, Cox JS, Welch JT et al. Pyrazinamide inhibits the eukaryotic-like fatty acid synthetase I (FASI) of Mycobacterium tuberculosis. *Nat Med* 2000; **6**: 1043-7.

334. Shi W, Zhang X, Jiang X et al. Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis. *Science (New York, NY* 2011; **333**: 1630-2.

335. A controlled trial of six months chemotherapy in pulmonary tuberculosis. Second report: results during the 24 months after the end of chemotherapy. British Thoracic Association. *Am Rev Respir Dis* 1982; **126**: 460-2.

336. Scorpio A, Zhang Y. Mutations in pncA, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. *Nat Med* 1996; **2**: 662-7.

337. Alexander DC, Ma JH, Guthrie JL et al. Gene Sequencing for Routine Verification of Pyrazinamide Resistance in Mycobacterium tuberculosis: a Role for pncA but Not rpsA. *J Clin Microbiol* 2012; **50**: 3726-8.

338. Barco P, Cardoso RF, Hirata RD et al. pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis clinical isolates from the southeast region of Brazil. *J Antimicrob Chemother* 2006; **58**: 930-5.

339. Bishop KS, Blumberg L, Trollip AP et al. Characterisation of the pncA gene in Mycobacterium tuberculosis isolates from Gauteng, South Africa. *Int J Tuberc Lung Dis* 2001; **5**: 952-7.

340. Cheng SJ, Thibert L, Sanchez T et al. pncA mutations as a major mechanism of pyrazinamide resistance in Mycobacterium tuberculosis: spread of a monoresistant strain in Quebec, Canada. *Antimicrob Agents Chemother* 2000; **44**: 528-32.

341. Chiu YC, Huang SF, Yu KW et al. Characteristics of pncA mutations in multidrug-resistant tuberculosis in Taiwan. *BMC Infect Dis* 2011; **11**: 240.

342. Hou L, Osei-Hyiaman D, Zhang Z et al. Molecular characterization of pncA gene mutations in Mycobacterium tuberculosis clinical isolates from China. *Epidemiol Infect* 2000; **124**: 227-32.

343. Jureen P, Werngren J, Toro JC et al. Pyrazinamide resistance and pncA gene mutations in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2008; **52**: 1852-4.

344. Kim HJ, Kwak HK, Lee J et al. Patterns of pncA mutations in drug-resistant Mycobacterium tuberculosis isolated from patients in South Korea. *Int J Tuberc Lung Dis* 2012; **16**: 98-103.

345. Lee KW, Lee JM, Jung KS. Characterization of pncA mutations of pyrazinamide-resistant Mycobacterium tuberculosis in Korea. *J Korean Med Sci* 2001; **16**: 537-43.

346. Lemaitre N, Sougakoff W, Truffot-Pernot C et al. Characterization of new mutations in pyrazinamide-resistant strains of Mycobacterium tuberculosis and identification of conserved regions important for the catalytic activity of the pyrazinamidase PncA. *Antimicrob Agents Chemother* 1999; **43**: 1761-3.

347. Mestdagh M, Fonteyne PA, Realini L et al. Relationship between pyrazinamide resistance, loss of pyrazinamidase activity, and mutations in the pncA locus in multidrug-resistant clinical isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 1999; **43**: 2317-9.

348. Morlock GP, Crawford JT, Butler WR et al. Phenotypic characterization of pncA mutants of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2000; **44**: 2291-5.

349. Muthaiah M, Jagadeesan S, Ayalusamy N et al. Molecular Epidemiological Study of Pyrazinamide-Resistance in Clinical Isolates of Mycobacterium tuberculosis from South India. *Int J Mol Sci* 2010; **11**: 2670-80.

350. Rodrigues Vde F, Telles MA, Ribeiro MO et al. Characterization of pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis in Brazil. *Antimicrob Agents Chemother* 2005; **49**: 444-6.

351. Somoskovi A, Dormandy J, Parsons LM et al. Sequencing of the pncA gene in members of the Mycobacterium tuberculosis complex has important diagnostic applications: Identification of a species-specific pncA mutation in "Mycobacterium canettii" and the reliable and rapid predictor of pyrazinamide resistance. *J Clin Microbiol* 2007; **45**: 595-9.

352. Sreevatsan S, Pan X, Zhang Y et al. Mutations associated with pyrazinamide resistance in pncA of Mycobacterium tuberculosis complex organisms. *Antimicrob Agents Chemother* 1997; **41**: 636-40.

353. Stoffels K, Mathys V, Fauville-Dufaux M et al. Systematic Analysis of Pyrazinamide-Resistant Spontaneous Mutants and Clinical Isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2012; **56**: 5186-93.

354. Tracevska T, Jansone I, Baumanis V et al. Spectrum of pncA mutations in multidrug-resistant Mycobacterium tuberculosis isolates obtained in Latvia. *Antimicrob Agents Chemother* 2004; **48**: 3209-10.

355. David HL. Probability distribution of drug-resistant mutants in unselected populations of Mycobacterium tuberculosis. *Appl Microbiol* 1970; **20**: 810-4.

356. Zimic M, Sheen P, Quiliano M et al. Peruvian and globally reported amino acid substitutions on the Mycobacterium tuberculosis pyrazinamidase suggest a conserved pattern of mutations associated to pyrazinamide resistance. *Infect Genet Evol* 2010; **10**: 346-9.

357. Unissa AN, Selvakumar N, Hassan S. Insight to pyrazinamide resistance in Mycobacterium tuberculosis by molecular docking. *Bioinformation* 2010; **4**: 24-9.

358. Felnagle EA, Podevels AM, Barkei JJ et al. Mechanistically distinct nonribosomal peptide synthetases assemble the structurally related antibiotics viomycin and capreomycin. *Chembiochem* 2011; **12**: 1859-67.

359. Maus CE, Plikaytis BB, Shinnick TM. Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2005; **49**: 3192-7.

360. Heifets L, Simon J, Pham V. Capreomycin is active against non-replicating M. tuberculosis. *Ann Clin Microbiol Antimicrob* 2005; **4**: 6.

361. Kondo J, Francois B, Russell RJ et al. Crystal structure of the bacterial ribosomal decoding site complexed with amikacin containing the gamma-amino-alpha-hydroxybutyryl (haba) group. *Biochimie* 2006; **88**: 1027-31.

362. Modolell J, Vazquez. The inhibition of ribosomal translocation by viomycin. *Eur J Biochem* 1977; **81**: 491-7.

363. Stanley RE, Blaha G, Grodzicki RL et al. The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat Struct Mol Biol* 2010; **17**: 289-93.

364. Yamada T, Bierhaus KH. Viomycin favours the formation of 70S ribosome couples. *Mol Gen Genet* 1978; **161**: 261-5.

365. Johansen SK, Maus CE, Plikaytis BB et al. Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs. *Mol Cell* 2006; **23**: 173-82.

366. Engstrom A, Morcillo N, Imperiale B et al. Detection of first- and second-line drug resistance in Mycobacterium tuberculosis clinical isolates by pyrosequencing. *J Clin Microbiol* 2012; **50**: 2026-33.

367. Feuerriegel S, Cox HS, Zarkua N et al. Sequence analyses of just four genes to detect extensively drug-resistant Mycobacterium tuberculosis strains in multidrug-resistant tuberculosis patients undergoing treatment. *Antimicrob Agents Chemother* 2009; **53**: 3353-6.

368. Jugheli L, Bzekalava N, de Rijk P et al. High level of cross-resistance between kanamycin, amikacin, and capreomycin among Mycobacterium tuberculosis isolates from Georgia and a close relation with mutations in the rrs gene. *Antimicrob Agents Chemother* 2009; **53**: 5064-8.

369. Via LE, Cho SN, Hwang S et al. Polymorphisms associated with resistance and cross-resistance to aminoglycosides and capreomycin in Mycobacterium tuberculosis isolates from South Korean Patients with drug-resistant tuberculosis. *J Clin Microbiol* 2010; **48**: 402-11.

370. Georghiou SB, Magana M, Garfein RS et al. Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to amikacin, kanamycin and capreomycin: a systematic review. *PLoS ONE* 2012; **7**: e33275.

371. Chen W, Biswas T, Porter VR et al. Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB. *Proc Natl Acad Sci U S A* 2011; **108**: 9804-8.

372. Zaunbrecher MA, Sikes RD, Jr., Metchock B et al. Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2009; **106**: 20004-9.

373. Engstrom A, Perskvist N, Werngren J et al. Comparison of clinical isolates and in vitro selected mutants reveals that tlyA is not a sensitive genetic marker for capreomycin resistance in Mycobacterium tuberculosis. *J Antimicrob Chemother* 2011; **66**: 1247-54.

374. Gikalo MB, Nosova EY, Krylova LY et al. The role of eis mutations in the development of kanamycin resistance in Mycobacterium tuberculosis isolates from the Moscow region. *J Antimicrob Chemother* 2012; **67**: 2107-9.

375. Maus CE, Plikaytis BB, Shinnick TM. Mutation of tlyA confers capreomycin resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2005; **49**: 571-7.

376. Monshupanee T, Johansen SK, Dahlberg AE et al. Capreomycin susceptibility is increased by TlyA-directed 2'-O-methylation on both ribosomal subunits. *Mol Microbiol* 2012; **85**: 1194-203.

377. Badet B, Hughes P, Kohiyama M et al. Inhibition of DNA replication in vitro by pefloxacin. *FEBS Lett* 1982; **145**: 355-9.

378. Malik S, Willby M, Sikes D et al. New insights into fluoroquinolone resistance in Mycobacterium tuberculosis: functional genetic analysis of gyrA and gyrB mutations. *PLoS ONE* 2012; **7**: e39754.

379. Piton J, Petrella S, Delarue M et al. Structural insights into the quinolone resistance mechanism of Mycobacterium tuberculosis DNA gyrase. *PLoS ONE* 2010; **5**: e12245.

380. Devasia R, Blackman A, Eden S et al. High proportion of fluoroquinolone-resistant Mycobacterium tuberculosis isolates with novel gyrase polymorphisms and a gyrA region associated with fluoroquinolone susceptibility. *J Clin Microbiol* 2012; **50**: 1390-6.

381. Migliori GB, Langendam MW, D'Ambrosio L et al. Protecting the tuberculosis drug pipeline: stating the case for the rational use of fluoroquinolones. *Eur Respir J* 2012; **40**: 814-22.

382. Devasia RA, Blackman A, Gebretsadik T et al. Fluoroquinolone resistance in Mycobacterium tuberculosis: the effect of duration and timing of fluoroquinolone exposure. *Am J Respir Crit Care Med* 2009; **180**: 365-70.

383. Malik M, Chavda K, Zhao X et al. Induction of mycobacterial resistance to quinolone class antimicrobials. *Antimicrob Agents Chemother* 2012; **56**: 3879-87.

384. Chen J, Chen Z, Li Y et al. Characterization of gyrA and gyrB mutations and fluoroquinolone resistance in Mycobacterium tuberculosis clinical isolates from Hubei Province, China. *Braz J Infect Dis* 2012; **16**: 136-41.

385. Von Groll A, Martin A, Jureen P et al. Fluoroquinolone resistance in Mycobacterium tuberculosis and mutations in gyrA and gyrB. *Antimicrob Agents Chemother* 2009; **53**: 4498-500.

386. Takiff HE, Salazar L, Guerrero C et al. Cloning and nucleotide sequence of Mycobacterium tuberculosis gyrA and gyrB genes and detection of quinolone resistance mutations. *Antimicrob Agents Chemother* 1994; **38**: 773-80.

387. Pantel A, Petrella S, Veziris N et al. Extending the definition of the GyrB quinolone resistance-determining region in Mycobacterium tuberculosis DNA gyrase for assessing fluoroquinolone resistance in M. tuberculosis. *Antimicrob Agents Chemother* 2012; **56**: 1990-6.

388. Cheng AF, Yew WW, Chan EW et al. Multiplex PCR amplimer conformation analysis for rapid detection of gyrA mutations in fluoroquinolone-resistant Mycobacterium tuberculosis clinical isolates. *Antimicrob Agents Chemother* 2004; **48**: 596-601.

389. Cui Z, Wang J, Lu J et al. Association of mutation patterns in gyrA/B genes and ofloxacin resistance levels in Mycobacterium tuberculosis isolates from East China in 2009. *BMC Infect Dis* 2011; **11**: 78.

390. Long Q, Li W, Du Q et al. gyrA/B fluoroquinolone resistance allele profiles amongst Mycobacterium tuberculosis isolates from mainland China. *Int J Antimicrob Agents* 2012; **39**: 486-9.

391. Suzuki Y, Nakajima C, Tamaru A et al. Sensitivities of ciprofloxacin-resistant Mycobacterium tuberculosis clinical isolates to fluoroquinolones: role of mutant DNA gyrase subunits in drug resistance. *Int J Antimicrob Agents* 2012; **39**: 435-9.

392. Pantel A, Petrella S, Matrat S et al. DNA gyrase inhibition assays are necessary to demonstrate fluoroquinolone resistance secondary to gyrB mutations in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2011; **55**: 4524-9.

393. Singh M, Jadaun GP, Ramdas et al. Effect of efflux pump inhibitors on drug susceptibility of ofloxacin resistant Mycobacterium tuberculosis isolates. *Indian J Med Res* 2011; **133**: 535-40.

394. Narayanan S. Molecular epidemiology of tuberculosis. *Indian J Med Res* 2004; **120**: 233-47.

395. Mathema B, Kurepina NE, Bifani PJ et al. Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 2006; **19**: 658-85.

396. Alland D, Kalkut GE, Moss AR et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* 1994; **330**: 1710-6.

397. Small PM, Hopewell PC, Singh SP et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* 1994; **330**: 1703-9.

398. Genewein A, Telenti A, Bernasconi C et al. Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 1993; **342**: 841-4.

399. Tabet SR, Goldbaum GM, Hooton TM et al. Restriction fragment length polymorphism analysis detecting a community-based tuberculosis outbreak among persons infected with human immunodeficiency virus. *J Infect Dis* 1994; **169**: 189-92.

400. Kanduma E, McHugh TD, Gillespie SH. Molecular methods for Mycobacterium tuberculosis strain typing: a users guide. *J Appl Microbiol* 2003; **94**: 781-91.

401. van Embden JD, Cave MD, Crawford JT et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993; **31**: 406-9.

402. Thierry D, Cave MD, Eisenach KD et al. IS6110, an IS-like element of Mycobacterium tuberculosis complex. *Nucleic Acids Res* 1990; **18**: 188.

403. van Soolingen D, Hermans PW, de Haas PE et al. Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991; **29**: 2578-86.

404. Cave MD, Eisenach KD, Templeton G et al. Stability of DNA fingerprint pattern produced with IS6110 in strains of Mycobacterium tuberculosis. *J Clin Microbiol* 1994; **32**: 262-6.

405. Portugal I, Maia S, Moniz-Pereira J. Discrimination of multidrug-resistant Mycobacterium tuberculosis IS6110 fingerprint subclusters by rpoB gene mutation analysis. *J Clin Microbiol* 1999; **37**: 3022-4.

406. Kamerbeek J, Schouls L, Kolk A et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J Clin Microbiol* 1997; **35**: 907-14.

407. Kremer K, van Soolingen D, Frothingham R et al. Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. *J Clin Microbiol* 1999; **37**: 2607-18.

408. Bauer J, Andersen AB, Kremer K et al. Usefulness of spoligotyping To discriminate IS6110 low-copy-number Mycobacterium tuberculosis complex strains cultured in Denmark. *J Clin Microbiol* 1999; **37**: 2602-6.

409. Soini H, Pan X, Teeter L et al. Transmission dynamics and molecular characterization of Mycobacterium tuberculosis isolates with low copy numbers of IS6110. *J Clin Microbiol* 2001; **39**: 217-21.

410. Warren RM, Streicher EM, Sampson SL et al. Microevolution of the direct repeat region of Mycobacterium tuberculosis: implications for interpretation of spoligotyping data. *J Clin Microbiol* 2002; **40**: 4457-65.

411. Mazars E, Lesjean S, Banuls AL et al. High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. *Proc Natl Acad Sci U S A* 2001; **98**: 1901-6.

412. Supply P, Mazars E, Lesjean S et al. Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. *Mol Microbiol* 2000; **36**: 762-71.

413. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977; **74**: 560-4.

414. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977; **74**: 5463-7.

415. Zhang J, Chiodini R, Badr A et al. The impact of next-generation sequencing on genomics. *J Genet Genomics* 2011; **38**: 95-109.

416. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* 2011; **52**: 413-35.

417. Margulies M, Egholm M, Altman WE et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**: 376-80.

418. Koser CU, Holden MT, Ellington MJ et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 2012; **366**: 2267-75.

419. McKernan KJ, Peckham HE, Costa GL et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome research* 2009; **19**: 1527-41.

420. Niedringhaus TP, Milanova D, Kerby MB et al. Landscape of next-generation sequencing technologies. *Anal Chem* 2011; **83**: 4327-41.

421. Coupland P, Chandra T, Quail M et al. Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. *Biotechniques* 2012; **53**: 365-72.

422. Eid J, Fehr A, Gray J et al. Real-time DNA sequencing from single polymerase molecules. *Science (New York, NY* 2009; **323**: 133-8.

423. Harris TD, Buzby PR, Babcock H et al. Single-molecule DNA sequencing of a viral genome. *Science (New York, NY* 2008; **320**: 106-9.

424. Rothberg JM, Hinz W, Rearick TM et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011; **475**: 348-52.

425. Stoddart D, Heron AJ, Klingelhoefer J et al. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. *Nano Lett* 2010; **10**: 3633-7.

426. Niemann S, Koser CU, Gagneux S et al. Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. *PLoS ONE* 2009; **4**: e7407.

427. Ioerger TR, Koo S, No EG et al. Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* 2009; **4**: e7778.

428. Schurch AC, Kremer K, Kiers A et al. The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. *Infect Genet Evol* 2009.

429. Ford CB, Lin PL, Chase MR et al. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. *Nat Genet* 2011; **43**: 482-6.

430. Casali N, Nikolayevskyy V, Balabanova Y et al. Microevolution of extensively drug-resistant tuberculosis in Russia. *Genome research* 2012.

431. CDC. Emergence of Mycobacterium tuberculosis with Extensive Resistance to Second-Line Drugs Worldwide, 2000-2004. *Morbidity and Mortality Weekly Report* 2006; **55**: 301-5.

432. Manissero D, Fernandez de la Hoz K. Surveillance methods and case definition for extensively drug resistant TB (XDR-TB) and relevance to Europe: summary update. *Euro Surveill* 2006; **11**: E061103 1.

433. Smittipat N, Billamas P, Palittapongarnpim M et al. Polymorphism of variable-number tandem repeats at multiple loci in Mycobacterium tuberculosis. *J Clin Microbiol* 2005; **43**: 5034-43.

434. Perdigao J, Macedo R, Silva C et al. From multidrug-resistant to extensively drug-resistant tuberculosis in Lisbon, Portugal: the stepwise mode of resistance acquisition. *J Antimicrob Chemother* 2013; **68**: 27-33.

435. Wong SY, Lee JS, Kwak HK et al. Mutations in gidB confer low-level streptomycin resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2011; **55**: 2515-22.

436. Direcção Geral de Saúde. Ponto da situação epidemiológica e indicadores de desempenho, Ano de 2004. Lisbon, Portugal. Direcção Geral de Saúde. 2005.

437. Cardoso RF, Cooksey RC, Morlock GP et al. Screening and characterization of mutations in isoniazid-resistant Mycobacterium tuberculosis isolates obtained in Brazil. *Antimicrob Agents Chemother* 2004; **48**: 3373-81.

438. Somoskovi A, Parsons LM, Salfinger M. The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in Mycobacterium tuberculosis. *Respir Res* 2001; **2**: 164-8.

439. Williams DL, Waguespack C, Eisenach K et al. Characterization of rifampin-resistance in pathogenic mycobacteria. *Antimicrob Agents Chemother* 1994; **38**: 2380-6.

440. Coll P. [Drugs with activity against Mycobacterium tuberculosis]. *Enferm Infecc Microbiol Clin* 2003; **21**: 299-307; quiz 8, 26.

441. Portugal I, Barreiro L, Moniz-Pereira J et al. pncA mutations in pyrazinamide-resistant Mycobacterium tuberculosis isolates in Portugal. *Antimicrob Agents Chemother* 2004; **48**: 2736-8.

442. Warren R, Richardson M, Sampson S et al. Genotyping of Mycobacterium tuberculosis with additional markers enhances accuracy in epidemiological studies. *J Clin Microbiol* 1996; **34**: 2219-24.

443. Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. *MMWR Morb Mortal Wkly Rep* 2006; **55**: 301-5.

444. Extensively drug-resistant tuberculosis (XDR-TB): recommendations for prevention and control. *Wkly Epidemiol Rec* 2006; **81**: 430-2.

445. van Soolingen D, de Haas PEW, Kremer K. Restriction fragment length polymorphism (RFLP) typing of mycobacteria. *Bilthoven, The Netherlands: National Intitute of Public Health and The Environment* 2002: 52.

446. Pfyffer GE, Bonato DA, Ebrahimzadeh A et al. Multicenter laboratory validation of susceptibility testing of Mycobacterium tuberculosis against classical second-line and newer antimicrobial drugs by using the radiometric BACTEC 460 technique and the proportion method with solid media. *J Clin Microbiol* 1999; **37**: 3179-86.

447. Uhl JR, Sandhu GS, Kline BC et al. PCR-RFLP detection of point mutations in the catalase-peroxidase gene (katG) of Mycobacterium tuberculosis associated with isoniazid resistance. *PCR protocols for emerging infectious disease ASM Press, Washington, DC*, 1996; 144–9.

448. Coll P, Aragon LM, Alcaide F et al. Molecular analysis of isoniazid and rifampin resistance in Mycobacterium tuberculosis isolates recovered from Barcelona. *Microb Drug Resist* 2005; **11**: 107-14.

449. Hannan MM, Desmond EP, Morlock GP et al. Pyrazinamide-monoresistant Mycobacterium tuberculosis in the United States. *J Clin Microbiol* 2001; **39**: 647-50.

450. Kirschner P, Bottger EC. Detection of mycobacterium resistance to streptomycin and clarithromycin. In: D.H. Persing, ed. *PCR protocols for emerging infectious diseases*. Washington, D.C.: American Society for Microbiology, 1996; 130–7.

451. Van Rie A, Warren R, Mshanga I et al. Analysis for a limited number of gene codons can predict drug resistance of Mycobacterium tuberculosis in a high-incidence community. *J Clin Microbiol* 2001; **39**: 636-41.

452. Direcção Geral de Saúde. Ponto da situação epidemiológica e indicadores de desempenho, Ano de 2003. Lisbon, Portugal. Direcção Geral de Saúde. 2004.

453. EuroTB and the national coordinators for tuberculosis surveillance in the WHO European Region. Report on tuberculosis cases notified in 2003. Saint-Maurice, France. Institut de veille sanitaire. 2005.

454. Cooksey RC, Morlock GP, Holloway BP et al. Temperature-mediated heteroduplex analysis performed by using denaturing high-performance liquid chromatography to identify sequence polymorphisms in Mycobacterium tuberculosis complex organisms. *J Clin Microbiol* 2002; **40**: 1610-6.

455. Lee AS, Lim IH, Tang LL et al. Contribution of kasA analysis to detection of isoniazid-resistant Mycobacterium tuberculosis in Singapore. *Antimicrob Agents Chemother* 1999; **43**: 2087-9.

456. Ramaswamy S, Musser JM. Molecular genetic basis of antimicrobial agent resistance in Mycobacterium tuberculosis: 1998 update. *Tuber Lung Dis* 1998; **79**: 3-29.

457. Huitric E, Werngren J, Jureen P et al. Resistance levels and rpoB gene mutations among in vitro-selected rifampin-resistant Mycobacterium tuberculosis mutants. *Antimicrob Agents Chemother* 2006; **50**: 2860-2.

458. Ma X, Wang H, Deng Y et al. rpoB Gene mutations and molecular characterization of rifampin-resistant Mycobacterium tuberculosis isolates from Shandong Province, China. *J Clin Microbiol* 2006; **44**: 3409-12.

459. Fukuda M, Koga H, Ohno H et al. Relationship between genetic alteration of the rpsL gene and streptomycin susceptibility of Mycobacterium tuberculosis in Japan. *J Antimicrob Chemother* 1999; **43**: 281-4.

460. Aragon LM, Navarro F, Heiser V et al. Rapid detection of specific gene mutations associated with isoniazid or rifampicin resistance in Mycobacterium tuberculosis clinical isolates using non-fluorescent low-density DNA microarrays. *J Antimicrob Chemother* 2006; **57**: 825-31.

461. Denkin S, Volokhov D, Chizhikov V et al. Microarray-based pncA genotyping of pyrazinamide-resistant strains of Mycobacterium tuberculosis. *J Med Microbiol* 2005; **54**: 1127-31.

462. Jureen P, Engstrand L, Eriksson S et al. Rapid detection of rifampin resistance in Mycobacterium tuberculosis by Pyrosequencing technology. *J Clin Microbiol* 2006; **44**: 1925-9.

463. Mieskes KT, Rusch-Gerdes S, Truffot-Pernot C et al. Rapid, simple, and culture-independent detection of rpsL codon 43 mutations that are highly predictive of streptomycin resistance in Mycobacterium tuberculosis. *Am J Trop Med Hyg* 2000; **63**: 56-60.

464. Masjedi MR, Farnia P, Sorooch S et al. Extensively drug-resistant tuberculosis: 2 years of surveillance in Iran. *Clin Infect Dis* 2006; **43**: 841-7.

465. Ferech M, Coenen S, Malhotra-Kumar S et al. European Surveillance of Antimicrobial Consumption (ESAC): outpatient quinolone use in Europe. *J Antimicrob Chemother* 2006; **58**: 423-7.

466. World Health Organization. Report of the meeting of the WHO Global Task Force on XDR-TB : Geneva, Switzerland, 9-10 October 2006. Geneva. World Health Organization. 2007. WHO Publication No. WHO/HTM/TB/2007.375

467. Francis J. Curry National Tuberculosis Center and California Department of Public Health. *Drug-Resistant Tuberculosis: A Survival Guide for Clinicians, Second Edition*. San Francisco, CA: Francis J. Curry National Tuberculosis Center, 2008.

468. LoBue P. Extensively drug-resistant tuberculosis. *Curr Opin Infect Dis* 2009; **22**: 167-73.

469. European Centre for Disease Prevention and Control/WHO Regional, Office for Europe. Tuberculosis surveillance in Europe 2007. Stockholm. European Centre for Disease Prevention and Control. 2009.

470. Jugheli L, Bzekalava N, de Rijk P et al. High cross resistance between kanamycin, amikacin and capreomycin among Mycobacterium tuberculosis isolates from Georgia and a close relation with mutations in the rrs gene. *Antimicrob Agents Chemother* 2009 Sep 14, Epub.

471. Hillemann D, Rusch-Gerdes S, Richter E. Feasibility of the GenoType MTBDRsl assay for fluoroquinolone, amikacin-capreomycin, and ethambutol resistance testing of Mycobacterium tuberculosis strains and clinical specimens. *J Clin Microbiol* 2009; **47**: 1767-72.

472. World Health Organization. Multidrug and extensively drug-resistant TB (M/XDR-TB): 2010 Global Report on Surveillance and Response. Geneva. World Health Organization 2010. WHO Publication No. WHO/HTM/TB/2010.3

473. Keshavjee S, Gelmanova IY, Farmer PE et al. Treatment of extensively drug-resistant tuberculosis in Tomsk, Russia: a retrospective cohort study. *Lancet* 2008; **372**: 1403-9.

474. Kim HR, Hwang SS, Kim HJ et al. Impact of extensive drug resistance on treatment outcomes in non-HIV-infected patients with multidrug-resistant tuberculosis. *Clin Infect Dis* 2007; **45**: 1290-5.

475. Mitnick CD, Shin SS, Seung KJ et al. Comprehensive treatment of extensively drug-resistant tuberculosis. *N Engl J Med* 2008; **359**: 563-74.

476. Devasia RA, Blackman A, Gebretsadik T et al. Fluoroquinolone resistance in Mycobacterium tuberculosis: the effect of duration and timing of fluoroquinolone exposure. *Am J Respir Crit Care Med* 2009; **180**: 365-70.

477. Gandhi NR, Moll A, Sturm AW et al. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* 2006; **368**: 1575-80.

478. Portugal I, Maia S, Moniz-Pereira J. Discrimination of Multidrug-Resistant Mycobacterium tuberculosis IS6110 Fingerprint Subclusters by rpoB Gene Mutation Analysis. *J Clin Microbiol* 1999.

479. Liippo KK, Kulmala K, Tala EO. Focusing tuberculosis contact tracing by smear grading of index cases. *Am Rev Respir Dis* 1993; **148**: 235-6.

480. Rajpal S, Dhingra VK, Aggarwal JK. Sputum grading as predictor of treatment outcome in pulmonary tuberculosis. *Indian Journal of Tuberculosis* 2002; **49**: 139-42.

481. Resch SC, Salomon JA, Murray M et al. Cost-effectiveness of treating multidrug-resistant tuberculosis. *PLoS Med* 2006; **3**: e241.

482. Perdigao J, Macedo R, Silva C et al. Tuberculosis drug-resistance in Lisbon, Portugal: a 6-year overview. *Clin Microbiol Infect* 2011; **17**: 1397-402.

483. Rusch-Gerdes S, Pfyffer GE, Casal M et al. Multicenter laboratory validation of the BACTEC MGIT 960 technique for testing susceptibilities of Mycobacterium tuberculosis to classical second-line drugs and newer antimicrobials. *J Clin Microbiol* 2006; **44**: 688-92.

484. Martin A, Camacho M, Portaels F et al. Resazurin microtiter assay plate testing of Mycobacterium tuberculosis susceptibilities to second-line drugs: rapid, simple, and inexpensive method. *Antimicrob Agents Chemother* 2003; **47**: 3616-9.

485. Martin A, Paasch F, Docx S et al. Multicentre laboratory validation of the colorimetric redox indicator (CRI) assay for the rapid detection of extensively drug-resistant (XDR) Mycobacterium tuberculosis. *J Antimicrob Chemother* 2011; **66**: 827-33.

486. Mitchison D, Davies G. The chemotherapy of tuberculosis: past, present and future. *Int J Tuberc Lung Dis* 2012; **16**: 724-32.

487. Beckert P, Hillemann D, Kohl TA et al. rplC T460C identified as a dominant mutation in linezolid-resistant Mycobacterium tuberculosis strains. *Antimicrob Agents Chemother* 2012; **56**: 2743-5.

488. Richter E, Rusch-Gerdes S, Hillemann D. First linezolid-resistant clinical isolates of Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2007; **51**: 1534-6.

489. Kim KH, An DR, Song J et al. Mycobacterium tuberculosis Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7. *Proc Natl Acad Sci U S A* 2012; **109**: 7729-34.

490. Menzies D, Benedetti A, Paydar A et al. Standardized treatment of active tuberculosis in patients with previous treatment and/or with mono-resistance to isoniazid: a systematic review and meta-analysis. *PLoS Med* 2009; **6**: e1000150.

491. European Centre for Disease Prevention and Control/WHO Regional Office for Europe. Tuberculosis surveillance in Europe 2008. Stockholm. European Centre for Disease Prevention and Control. 2010.

492. Meier A, Kirschner P, Bange FC et al. Genetic alterations in streptomycin-resistant Mycobacterium tuberculosis: mapping of mutations conferring resistance. *Antimicrobial Agents and Chemotherapy* 1994; **38**: 228-33.

493. Mikheil DM, Shippy DC, Eakley NM et al. Deletion of gene encoding methyltransferase (gidB) confers high-level antimicrobial resistance in Salmonella. *J Antibiot (Tokyo)* 2012; **65**: 185-92.

494. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073-81.

495. Romanowski MJ, Bonanno JB, Burley SK. Crystal structure of the Escherichia coli glucose-inhibited division protein B (GidB) reveals a methyltransferase fold. *Proteins* 2002; **47**: 563-7.

496. Kumar A, Saigal K, Malhotra K et al. Structural and functional characterization of Rv2966c protein reveals an RsmD-like methyltransferase from Mycobacterium tuberculosis and the role of its N-terminal domain in target recognition. *J Biol Chem* 2011; **286**: 19652-61.

497. Nhu NT, Lan NT, Phuong NT et al. Association of streptomycin resistance mutations with level of drug resistance and Mycobacterium tuberculosis genotypes. *Int J Tuberc Lung Dis* 2012; **16**: 527-31.

498. David S, Ribeiro DR, Antunes A et al. Contribution of spoligotyping to the characterization of the population structure of Mycobacterium tuberculosis isolates in Portugal. *Infect Genet Evol* 2007; **7**: 609-17.

499. Sharma SK, Kumar S, Saha PK et al. Prevalence of multidrug-resistant tuberculosis among category II pulmonary tuberculosis patients. *Indian J Med Res* 2011; **133**: 312-5.

500. Ponce M, Ugarte-Gil C, Zamudio C et al. Additional evidence to support the phasing-out of treatment category II regimen for pulmonary tuberculosis in Peru. *Trans R Soc Trop Med Hyg* 2012; **106**: 508-10.

501. Furin J, Gegia M, Mitnick C et al. Eliminating the category II retreatment regimen from national tuberculosis programme guidelines: the Georgian experience. *Bull World Health Organ* 2012; **90**: 63-6.

502. Jones-Lopez EC, Ayakaka I, Levin J et al. Effectiveness of the standard WHO recommended retreatment regimen (category II) for tuberculosis in Kampala, Uganda: a prospective cohort study. *PLoS Med* 2011; **8**: e1000427.

503. Abubakar I, Zignol M, Falzon D et al. Drug-resistant tuberculosis: time for a visionary political leadership. *Lancet Infect Dis* 2013.

504. Ioerger TR, Feng Y, Ganesula K et al. Variation among genome sequences of H37Rv strains of Mycobacterium tuberculosis from multiple laboratories. *J Bacteriol* 2010; **192**: 3645-53.

505. Ford C, Yusim K, Ioerger T et al. Mycobacterium tuberculosis--heterogeneity revealed through whole genome sequencing. *Tuberculosis (Edinburgh, Scotland)* 2012; **92**: 194-201.

506. Machado D, Perdigao J, Ramos J et al. High-level resistance to isoniazid and ethionamide in multidrug-resistant Mycobacterium tuberculosis of the Lisboa family is associated with inhA double mutations. *J Antimicrob Chemother* 2013.

507. Comas I, Borrell S, Roetzer A et al. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012; **44**: 106-10.

508. de Vos M, Muller B, Borrell S et al. Putative compensatory mutations in the rpoC gene of rifampicin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. *Antimicrob Agents Chemother* 2012.

509. Lazzarini LC, Huard RC, Boechat NL et al. Discovery of a novel Mycobacterium tuberculosis lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. *J Clin Microbiol* 2007; **45**: 3891-902.

510. Gibson AL, Huard RC, Gey van Pittius NC et al. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean Mycobacterium tuberculosis spoligotype family. *J Clin Microbiol* 2008; **46**: 1259-67.

511. Madhavilatha GK, Joseph BV, Paul LK et al. Whole-genome sequences of two clinical isolates of Mycobacterium tuberculosis from Kerala, South India. *J Bacteriol* 2012; **194**: 4430.

512. Wong K, Keane TM, Stalker J et al. Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biol* 2010; **11**: R128.

513. Casart Y, Turcios L, Florez I et al. IS6110 in oriC affects the morphology and growth of Mycobacterium tuberculosis and attenuates virulence in mice. *Tuberculosis (Edinburgh, Scotland)* 2008; **88**: 545-52.

514. Kurepina N, Likhoshvay E, Shashkina E et al. Targeted hybridization of IS6110 fingerprints identifies the W-Beijing Mycobacterium tuberculosis strains among clinical isolates. *J Clin Microbiol* 2005; **43**: 2148-54.

515. Plikaytis BB, Marden JL, Crawford JT et al. Multiplex PCR assay specific for the multidrug-resistant strain W of Mycobacterium tuberculosis. *J Clin Microbiol* 1994; **32**: 1542-6.

516. Lin J, Sattar AN, Puckree T. An alarming rate of drug-resistant tuberculosis at Ngwelezane Hospital in northern KwaZulu Natal, South Africa. *Int J Tuberc Lung Dis* 2004; **8**: 568-73.

517. Pillay M, Sturm AW. Evolution of the extensively drug-resistant F15/LAM4/KZN strain of Mycobacterium tuberculosis in KwaZulu-Natal, South Africa. *Clin Infect Dis* 2007; **45**: 1409-14.

518. Lazzarini LC, Spindola SM, Bang H et al. RDRio Mycobacterium tuberculosis infection is associated with a higher frequency of cavitary pulmonary disease. *J Clin Microbiol* 2008; **46**: 2175-83.

519. Akhter Y, Ehebauer MT, Mukhopadhyay S et al. The PE/PPE multigene family codes for virulence factors and is a possible source of mycobacterial antigenic variation: perhaps more? *Biochimie* 2012; **94**: 110-6.

520. Mukhopadhyay S, Balaji KN. The PE and PPE proteins of Mycobacterium tuberculosis. *Tuberculosis (Edinburgh, Scotland)* 2011; **91**: 441-7.

521. Behr MA, Warren SA, Salamon H et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet* 1999; **353**: 444-9.

522. Hernandez-Garduno E, Cook V, Kunimoto D et al. Transmission of tuberculosis from smear negative patients: a molecular epidemiology study. *Thorax* 2004; **59**: 286-90.

523. Tostmann A, Kik SV, Kalisvaart NA et al. Tuberculosis transmission by patients with smear-negative pulmonary tuberculosis in a large cohort in the Netherlands. *Clin Infect Dis* 2008; **47**: 1135-42.

524. Choudhary RK, Mukhopadhyay S, Chakhaiyar P et al. PPE antigen Rv2430c of Mycobacterium tuberculosis induces a strong B-cell response. *Infect Immun* 2003; **71**: 6338-43.

525. Tundup S, Pathak N, Ramanadham M et al. The co-operonic PE25/PPE41 protein complex of Mycobacterium tuberculosis elicits increased humoral and cell mediated immune response. *PLoS ONE* 2008; **3**: e3586.

526. Fenner L, Egger M, Bodmer T et al. Effect of mutation and genetic background on drug resistance in Mycobacterium tuberculosis. *Antimicrob Agents Chemother* 2012; **56**: 3047-53.

527. Brimacombe M, Hazbon M, Motiwala AS et al. Antibiotic resistance and single-nucleotide polymorphism cluster grouping type in a multinational sample of resistant Mycobacterium tuberculosis isolates. *Antimicrob Agents Chemother* 2007; **51**: 4157-9.

528. Richardson ET, Lin SY, Pinsky BA et al. First documentation of isoniazid reversion in Mycobacterium tuberculosis. *Int J Tuberc Lung Dis* 2009; **13**: 1347-54.

529. Brandis G, Wrande M, Liljas L et al. Fitness-compensatory mutations in rifampicin-resistant RNA polymerase. *Mol Microbiol* 2012; **85**: 142-51.

530. Opalka N, Brown J, Lane WJ et al. Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. *PLoS Biol* 2010; **8**.

531. Sherman DR, Mdluli K, Hickey MJ et al. Compensatory ahpC gene expression in isoniazid-resistant Mycobacterium tuberculosis. *Science (New York, NY* 1996; **272**: 1641-3.

532. Shcherbakov D, Akbergenov R, Matt T et al. Directed mutagenesis of Mycobacterium smegmatis 16S rRNA to reconstruct the in-vivo evolution of aminoglycoside resistance in Mycobacterium tuberculosis. *Mol Microbiol* 2010.

533. Gagneux S, Burgos MV, DeRiemer K et al. Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. *PLoS Pathog* 2006; **2**: e61.

534. de Boer AS, Borgdorff MW, de Haas PE et al. Analysis of rate of change of IS6110 RFLP patterns of Mycobacterium tuberculosis based on serial patient isolates. *J Infect Dis* 1999; **180**: 1238-44.

535. Yeh RW, Ponce de Leon A, Agasino CB et al. Stability of Mycobacterium tuberculosis DNA genotypes. *J Infect Dis* 1998; **177**: 1107-11.

536. Sampson S, Warren R, Richardson M et al. IS6110 insertions in Mycobacterium tuberculosis: predominantly into coding regions. *J Clin Microbiol* 2001; **39**: 3423-4.

537. Tanaka MM, Rosenberg NA, Small PM. The control of copy number of IS6110 in Mycobacterium tuberculosis. *Mol Biol Evol* 2004; **21**: 2195-201.

538. Thorne N, Borrell S, Evans J et al. IS6110-based global phylogeny of Mycobacterium tuberculosis. *Infect Genet Evol* 2011; **11**: 132-8.

539. Beste DJ, Hooper T, Stewart G et al. GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosis metabolism. *Genome Biol* 2007; **8**: R89.

540. Fang X, Wallqvist A, Reifman J. Development and analysis of an in vivo-compatible metabolic network of Mycobacterium tuberculosis. *BMC Syst Biol* 2010; **4**: 160.

541. Jamshidi N, Palsson BO. Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol* 2007; **1**: 26.

542. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754-60.

543. Li H, Handsaker B, Wysoker A et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078-9.

544. Danecek P, Auton A, Abecasis G et al. The variant call format and VCFtools. *Bioinformatics* 2011; **27**: 2156-8.

545. Ye K, Schulz MH, Long Q et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009; **25**: 2865-71.

546. Chen K, Wallis JW, McLellan MD et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009; **6**: 677-81.

547. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 2008; **18**: 821-9.

548. Shriver MD, Jin L, Boerwinkle E et al. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol Biol Evol* 1995; **12**: 914-20.

549. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; **27**: 221-4.

550. Machado D, Perdigão J, Ramos J et al. High-level resistance to isoniazid and ethionamide in multidrug-resistant Mycobacterium tuberculosis of the Lisboa family is associated with inhA double mutations. *J Antimicrob Chemother* 2013: *in press*.

551. Supply P. Multilocus Variable Number Tandem Repeat. Genotyping of Mycobacterium tuberculosis. Technical Guide. Institut de Biologie/Institut Pasteur de Lille, 2005.

552. van Soolingen D, Schurch AC, Kremer K et al. Molecular Epidemiology; Professional Public Health Instrument or Research Tool? In: 32nd Annual Congress of the European Society of Mycobacteriology, ed. Lübeck, Germany, 2011.

553. Dahl JL, Wei J, Moulder JW et al. Subcellular localization of the Iitracellular survival-enhancing Eis protein of Mycobacterium tuberculosis. *Infect Immun* 2001; **69**: 4295-302.

554. Wei J, Dahl JL, Moulder JW et al. Identification of a Mycobacterium tuberculosis gene that enhances mycobacterial survival in macrophages. *J Bacteriol* 2000; **182**: 377-84.

555. Wu S, Barnes PF, Samten B et al. Activation of the eis gene in a W-Beijing strain of Mycobacterium tuberculosis correlates with increased SigA levels and enhanced intracellular growth. *Microbiology* 2009; **155**: 1272-81.

556. Wu S, Howard ST, Lakey DL et al. The principal sigma factor sigA mediates enhanced growth of Mycobacterium tuberculosis in vivo. *Mol Microbiol* 2004; **51**: 1551-62.

557. Munsiff SS, Nivin B, Sacajiu G et al. Persistence of a highly resistant strain of tuberculosis in New York City during 1990-1999. *J Infect Dis* 2003; **188**: 356-63.

558. Yoshiyama T, Yanai H, Rhiengtong D et al. Development of acquired drug resistance in recurrent tuberculosis patients with various previous treatment outcomes. *Int J Tuberc Lung Dis* 2004; **8**: 31-8.

559. Seddon JA, Warren RM, Enarson DA et al. Drug-resistant tuberculosis transmission and resistance amplification within families. *Emerg Infect Dis* 2012; **18**: 1342-5.

560. Cox HS, Niemann S, Ismailov G et al. Risk of acquired drug resistance during short-course directly observed treatment of tuberculosis in an area with high levels of drug resistance. *Clin Infect Dis* 2007; **44**: 1421-7.

561. Shin SS, Keshavjee S, Gelmanova IY et al. Development of extensively drug-resistant tuberculosis during multidrug-resistant tuberculosis treatment. *Am J Respir Crit Care Med* 2010; **182**: 426-32.

562. Matthys F, Rigouts L, Sizaire V et al. Outcomes after chemotherapy with WHO category II regimen in a population with high prevalence of drug resistant tuberculosis. *PLoS ONE* 2009; **4**: e7954.

563. Bonilla CA, Crossa A, Jave HO et al. Management of extensively drug-resistant tuberculosis in Peru: cure is possible. *PLoS One* 2008; **3**: e2957.

564. Veigas B, Machado D, Perdigao J et al. Au-nanoprobes for detection of SNPs associated with antibiotic resistance in Mycobacterium tuberculosis. *Nanotechnology* 2010; **21**: 415101.

565. Pfyffer GE, Strassle A, Rose N et al. Transmission of tuberculosis in the metropolitan area of Zurich: a 3 year survey based on DNA fingerprinting. *Eur Respir J* 1998; **11**: 804-8.

566. van Deutekom H, Gerritsen JJ, van Soolingen D et al. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. *Clin Infect Dis* 1997; **25**: 1071-7.

567. Frieden TR, Fujiwara PI, Washko RM et al. Tuberculosis in New York City--turning the tide. *N Engl J Med* 1995; **333**: 229-33.

568. da Costa PA, Trajman A, Mello FC et al. Administrative measures for preventing Mycobacterium tuberculosis infection among healthcare workers in a teaching hospital in Rio de Janeiro, Brazil. *J Hosp Infect* 2009; **72**: 57-64.

569. Corbett EL, Bandason T, Duong T et al. Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster-randomised trial. *Lancet* 2010; **376**: 1244-53.

570. Davies AP, Billington OJ, Bannister BA et al. Comparison of fitness of two isolates of Mycobacterium tuberculosis, one of which had developed multi-drug resistance during the course of treatment. *J Infect* 2000; **41**: 184-7.

571. Dye C, Espinal MA. Will tuberculosis become resistant to all antibiotics? *Proc Biol Sci* 2001; **268**: 45-52.

572. Cohen T, Murray M. Modeling epidemics of multidrug-resistant M. tuberculosis of heterogeneous fitness. *Nat Med* 2004; **10**: 1117-21.

573. Luciani F, Sisson SA, Jiang H et al. The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 2009; **106**: 14711-5.

574. Muller B, Borrell S, Rose G et al. The heterogeneous evolution of multidrug-resistant Mycobacterium tuberculosis. *Trends Genet* 2013; **29**: 160-9.

575. Sander P, Springer B, Prammananan T et al. Fitness cost of chromosomal drug resistance-conferring mutations. *Antimicrob Agents Chemother* 2002; **46**: 1204-11.

SUPPLEMENTARY MATERIAL

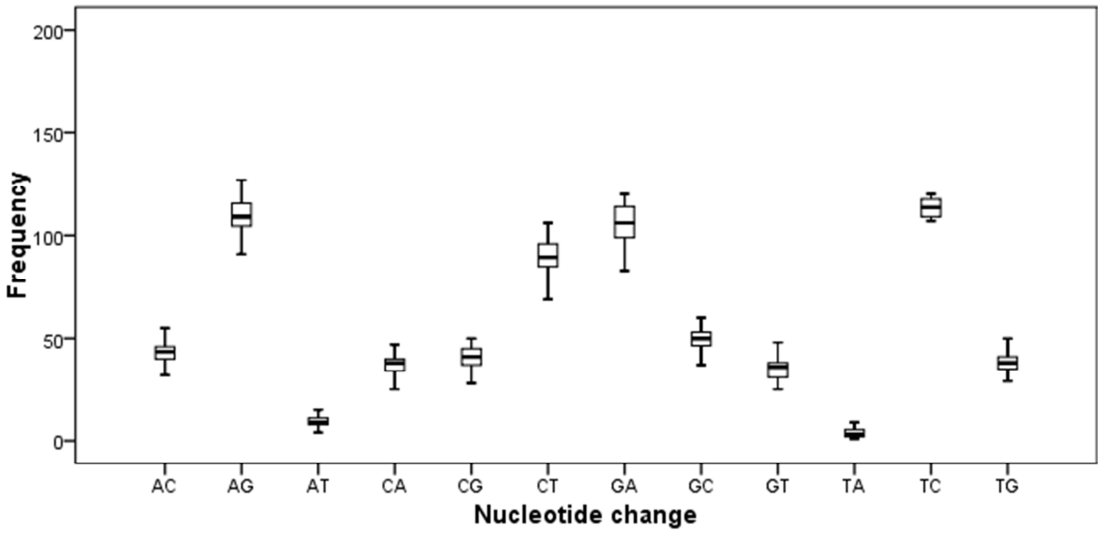


Figure S7.1 - Boxplot graph showing the different types of SNP mutations.

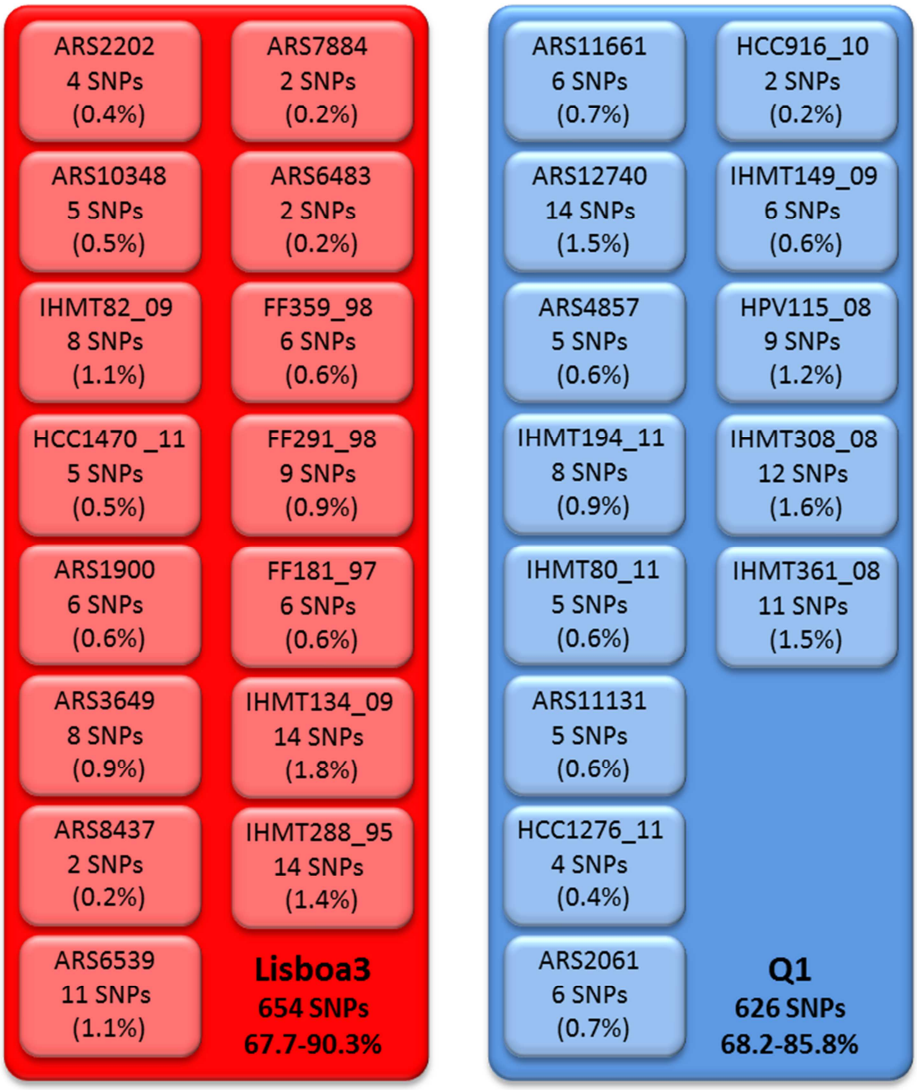


Figure S7.2 – Intra-clade SNP diversity and uniqueness. Number of SNPs unique to each isolate and percentage of total SNPs detected. Represented below each clade designation are: the number of SNPs that represents the total pool of SNPs shared by all isolates belonging to the respective clade; and, the range of the total percentage that this latter SNP pool count comprises from the total percentage of the isolates' detected SNPs.

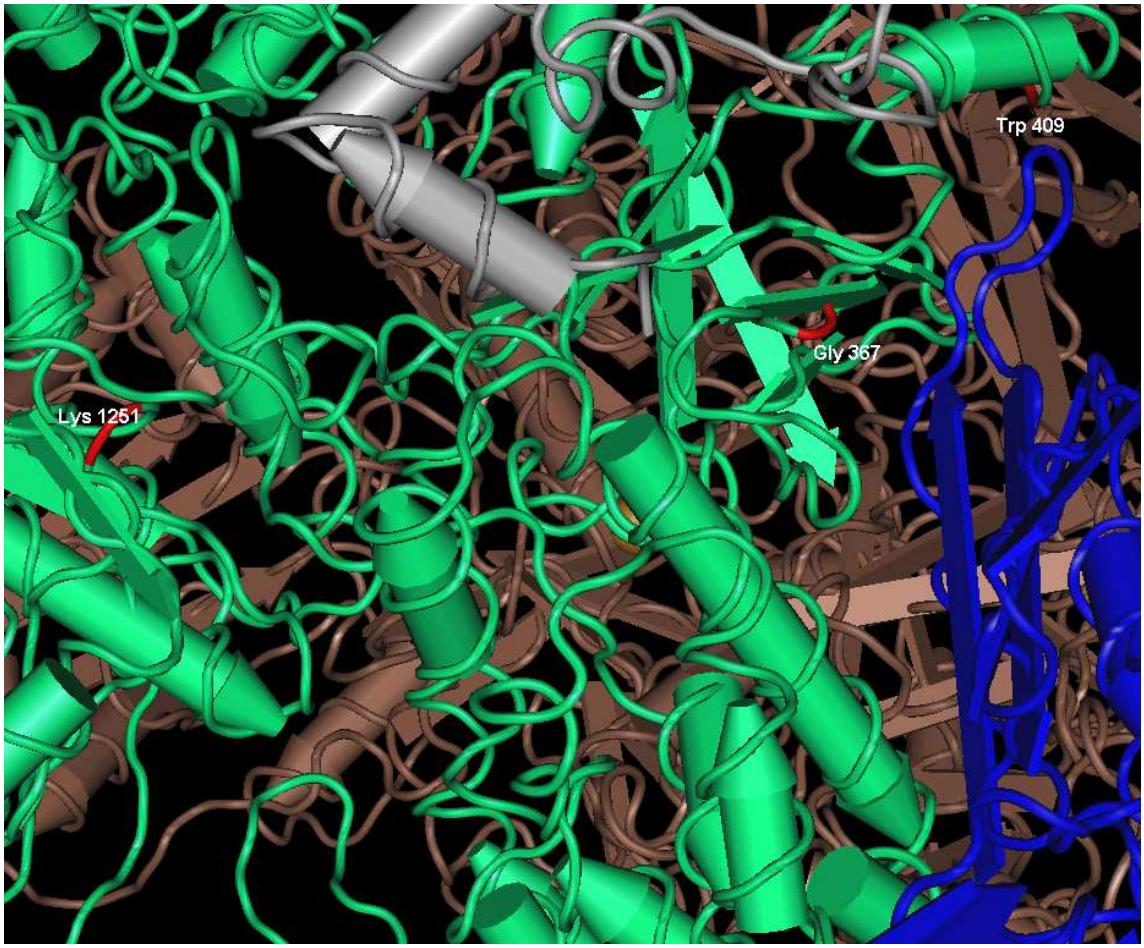


Figure S7.3 – Molecular model of Escherichia coli core RNA polymerase (Opalka et al. 2010) (RCSB Protein Data Bank ref. 3LU0) showing the homologous RpoC residues found to be involved in putative RIF resistance compensation in *M. tuberculosis*. The different RNA polymerase subunits are shown: Alpha/RpoA (blue chain), Beta/RpoB (brown chain), Beta'/RpoC (green chain) and Omega/RpoZ (grey chain). The RpoC highlighted residues, in red, Gly367, Trp409 and Lys1251 are homologous to the RpoC residues Gly442, Trp484 and Lys1152 from *M. tuberculosis*, respectively.

REFERENCES

Opalka N, Brown J, Lane WJ, Twist KA, Landick R, Asturias FJ, Darst SA. 2010. Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. *PLoS Biol* **8**(9).

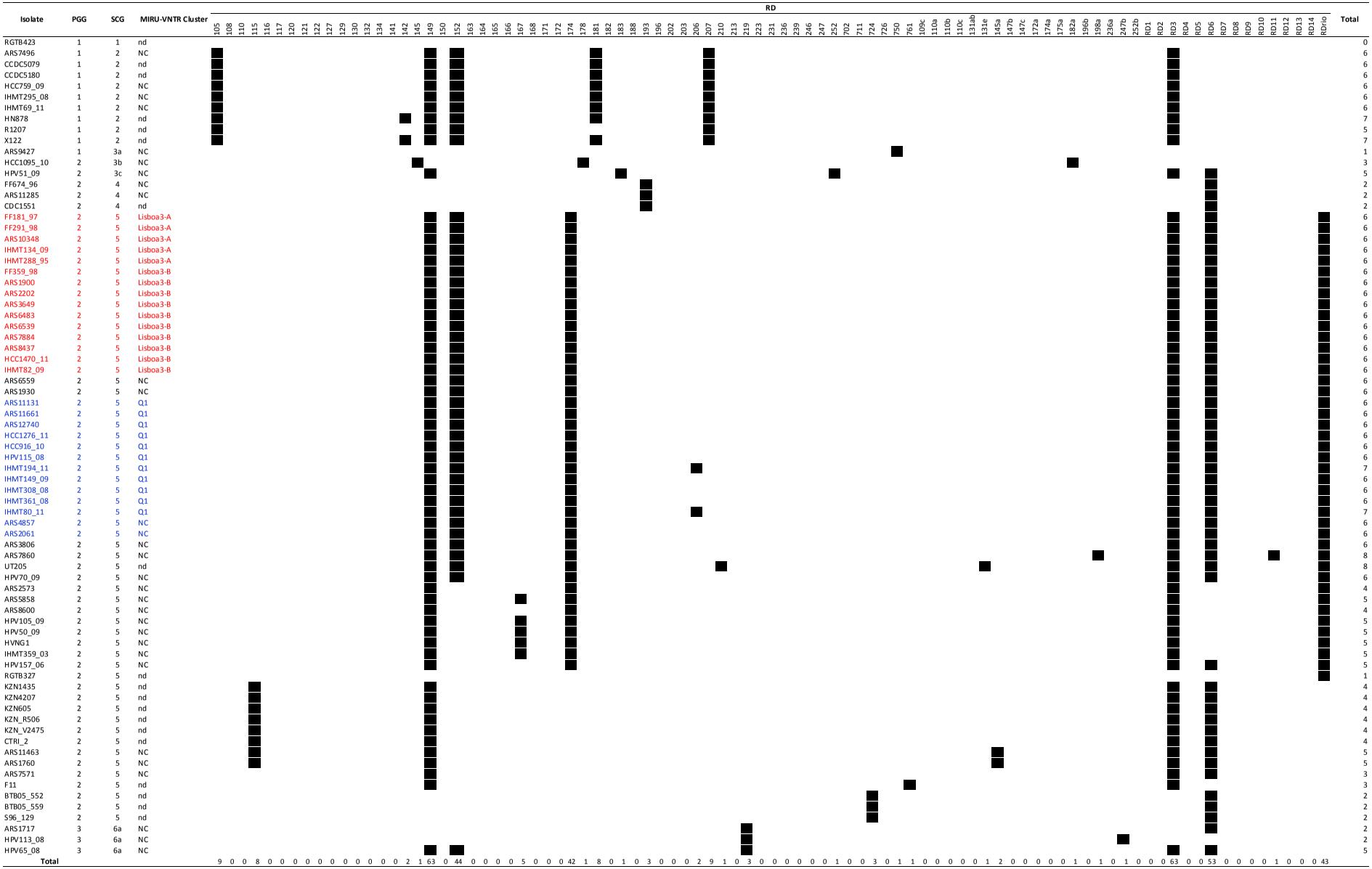
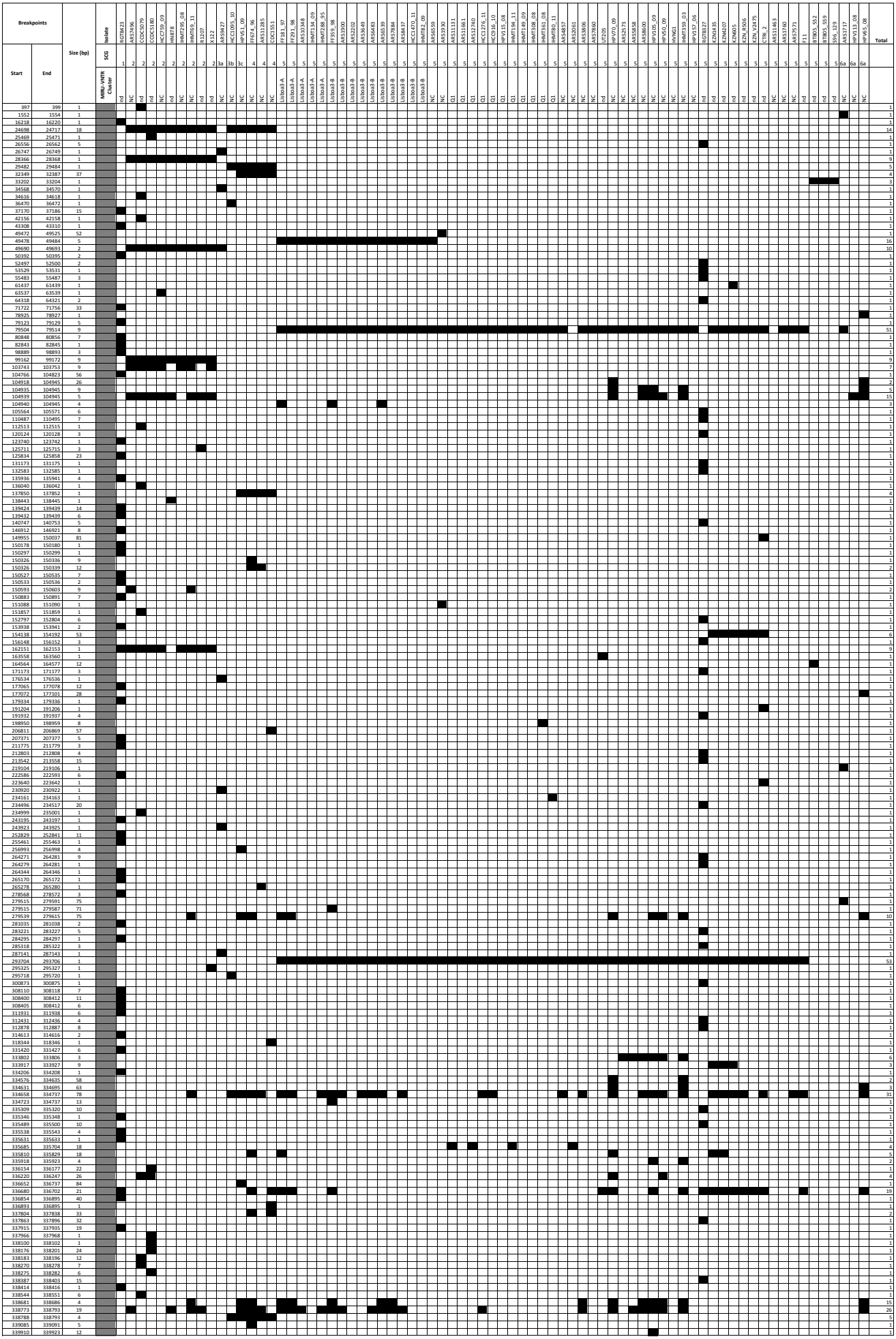
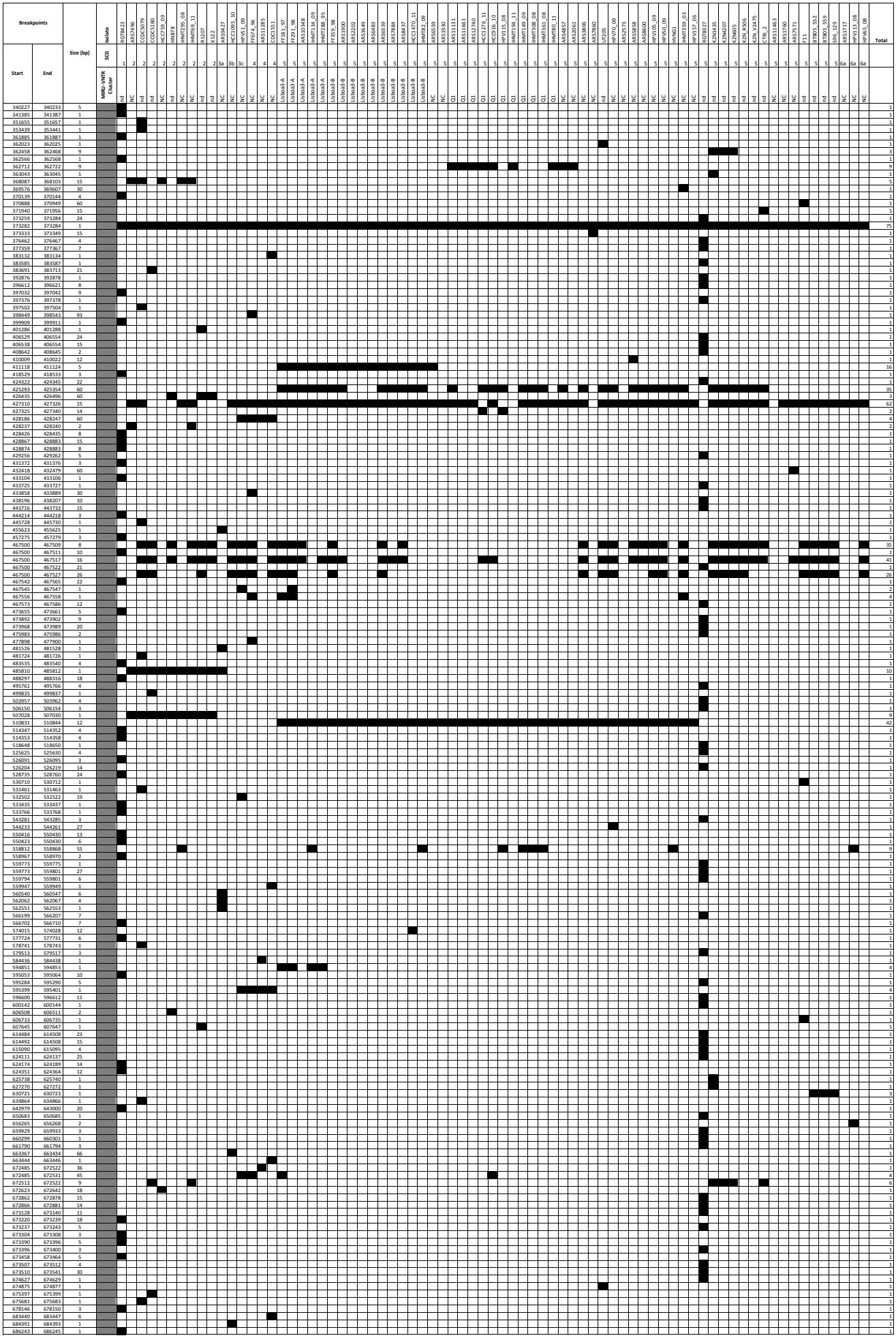
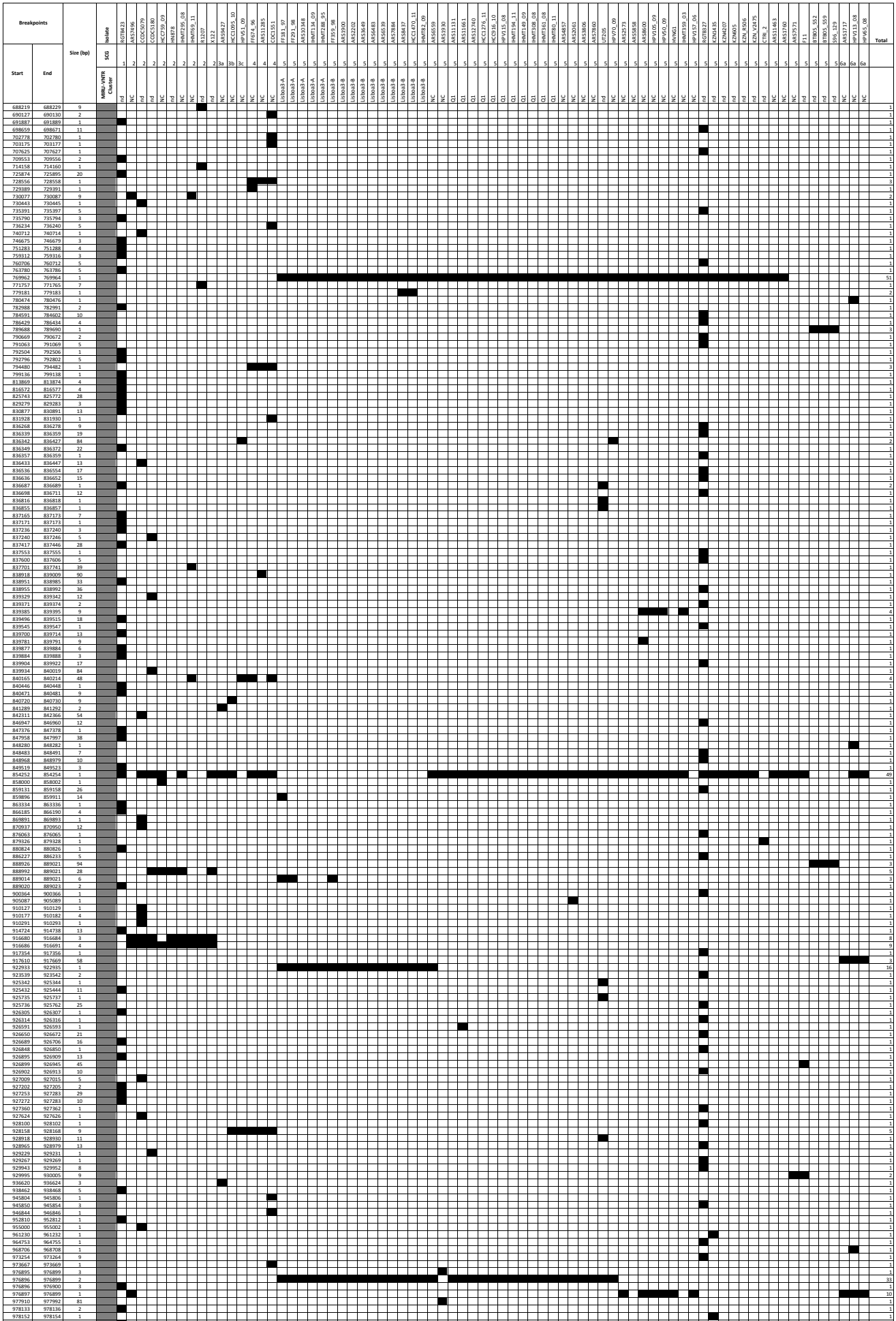
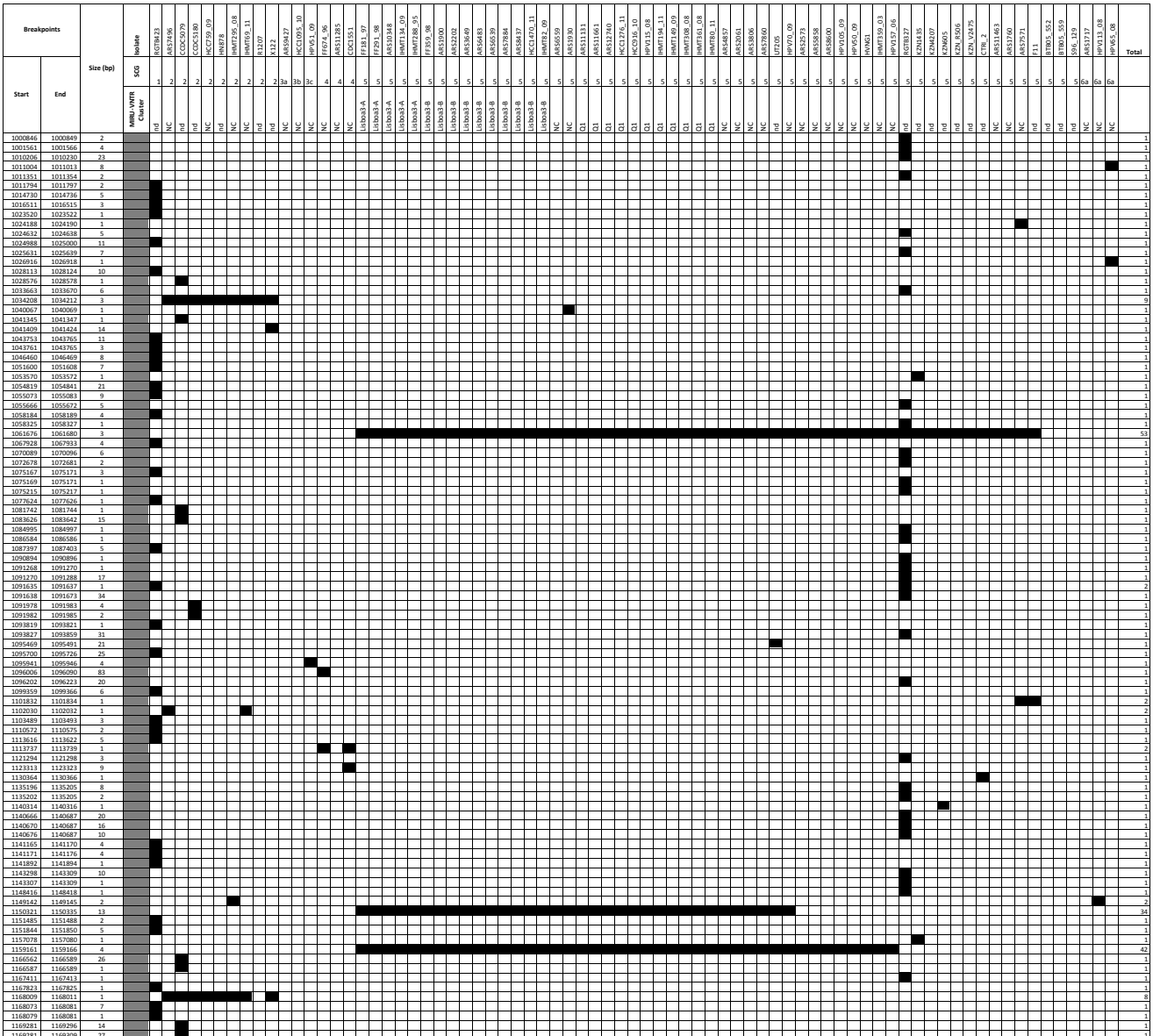


Figure S7.4 - Distribution of RD deletions found across the analyzed genomes of 75 *M. tuberculosis* isolates. RD absence is assigned with a black square and, Lisboa3 and Q1 clade isolates are highlighted in red and blue, respectively. Column and line totals account for the total number of RDs in column or line, respectively.

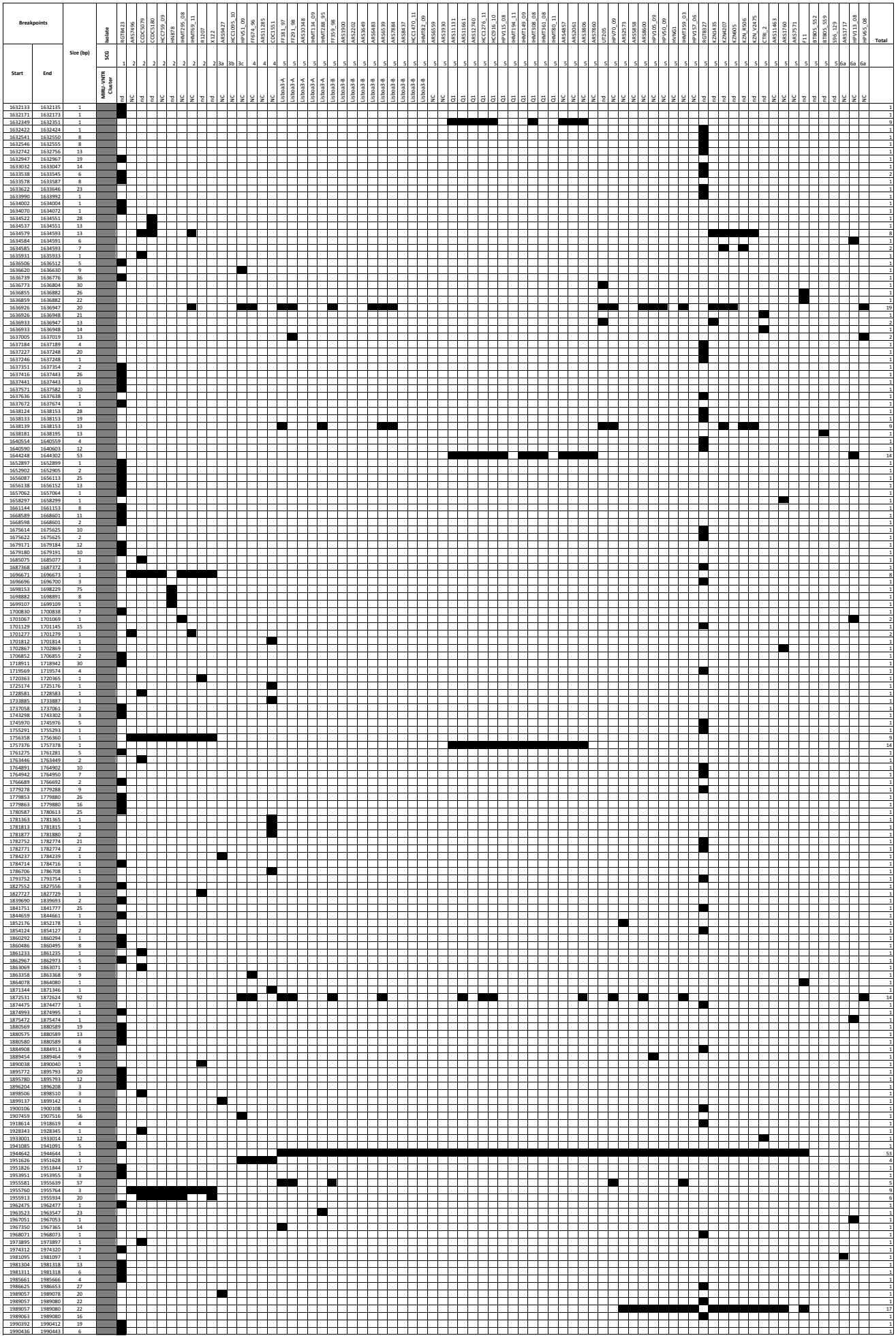

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NRRL-UNTR Cluster
Total

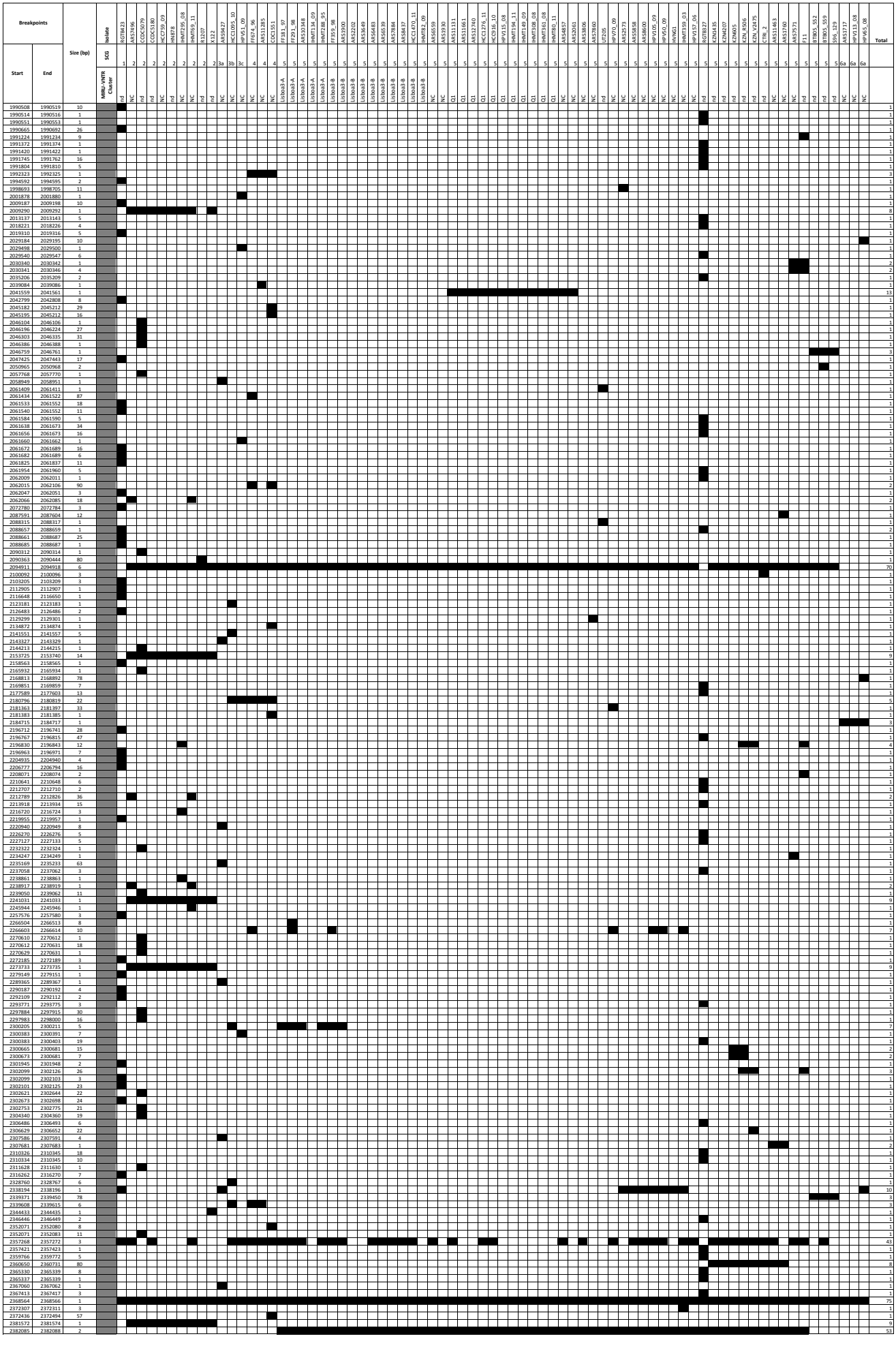

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NRMI-I/HTR Cluster
Total
1
2
3a
3b
3c
4
5
6a
6b
nd
NC
Q1

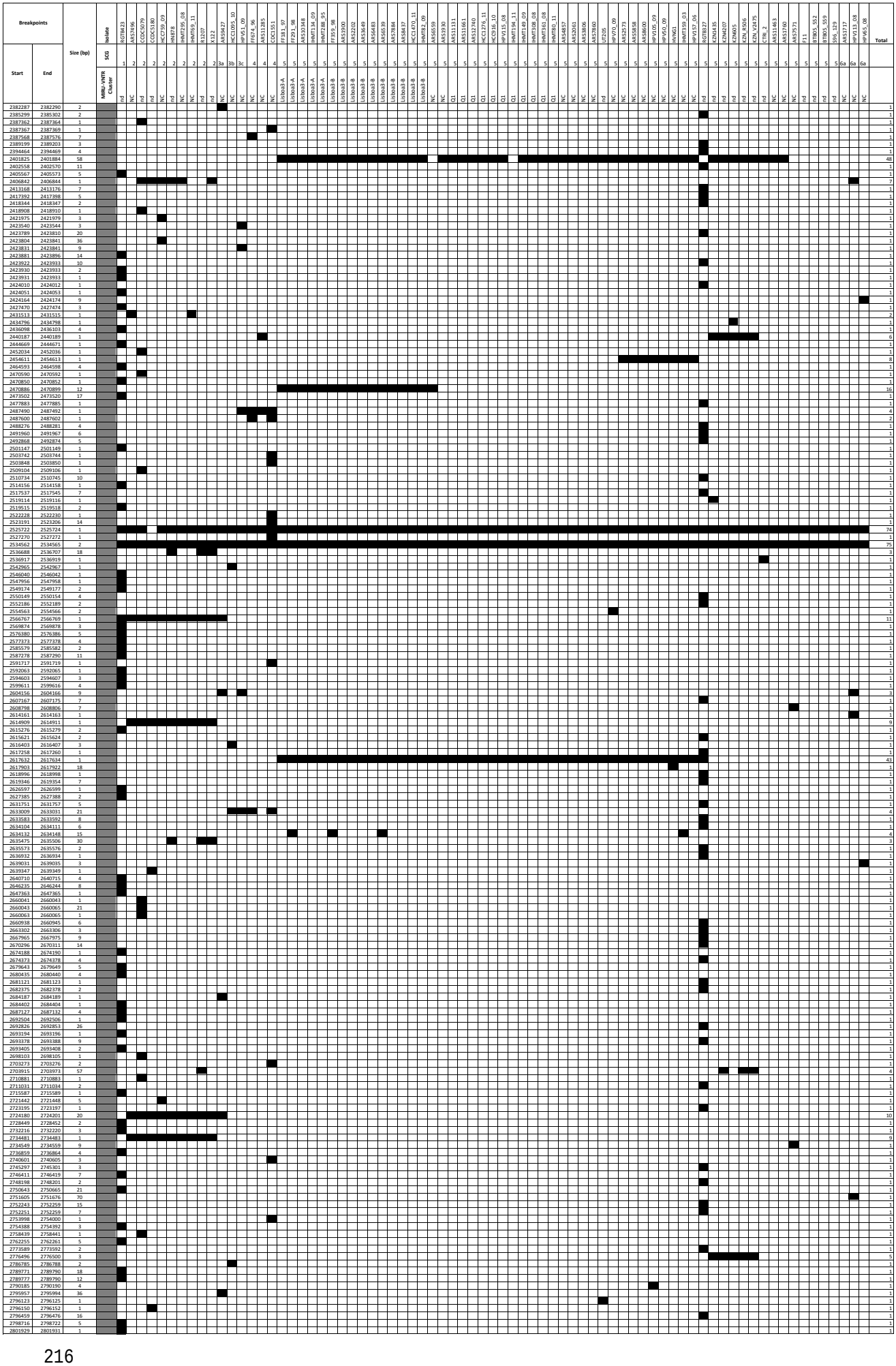

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NMNJ-LHVTR Cluster
ISOTH431
ABS7496
CDC6079
CDC5180
HCC739_09
HNI878
HMT295_08
HMT69_11
R1207
X122
ABS9427
HCC1095_10
HPV51_09
FF674_96
C0C1351
ABS11285
FF181_97
FF291_98
ABS91348
HMT134_09
HMT288_05
FF393_96
ABS1900
ABS2202
ABS36149
ABS6483
ABS65139
ABS7684
ABS8437
HCC1470_11
HMT82_09
ABS6559
ABS1930
ABS11131
ABS11661
ABS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT108_08
HMT161_08
HMT80_11
ABS4857
ABS2061
ABS3806
ABS7860
UTZ-05
HPV70_09
ABS2573
ABS3858
ABS8600
HPV105_09
HPV50_09
HVN61
HMT139_03
HPV157_06
ISOTH427
K2N4135
K2N4107
K2N605
K2N_K906
K2N_V2475
CTR_2
ABS11463
ABS3760
ABS791
F11
BTB03_552
BTB03_559
596_129
ABS1717
HPV113_08
HPV65_08
Total

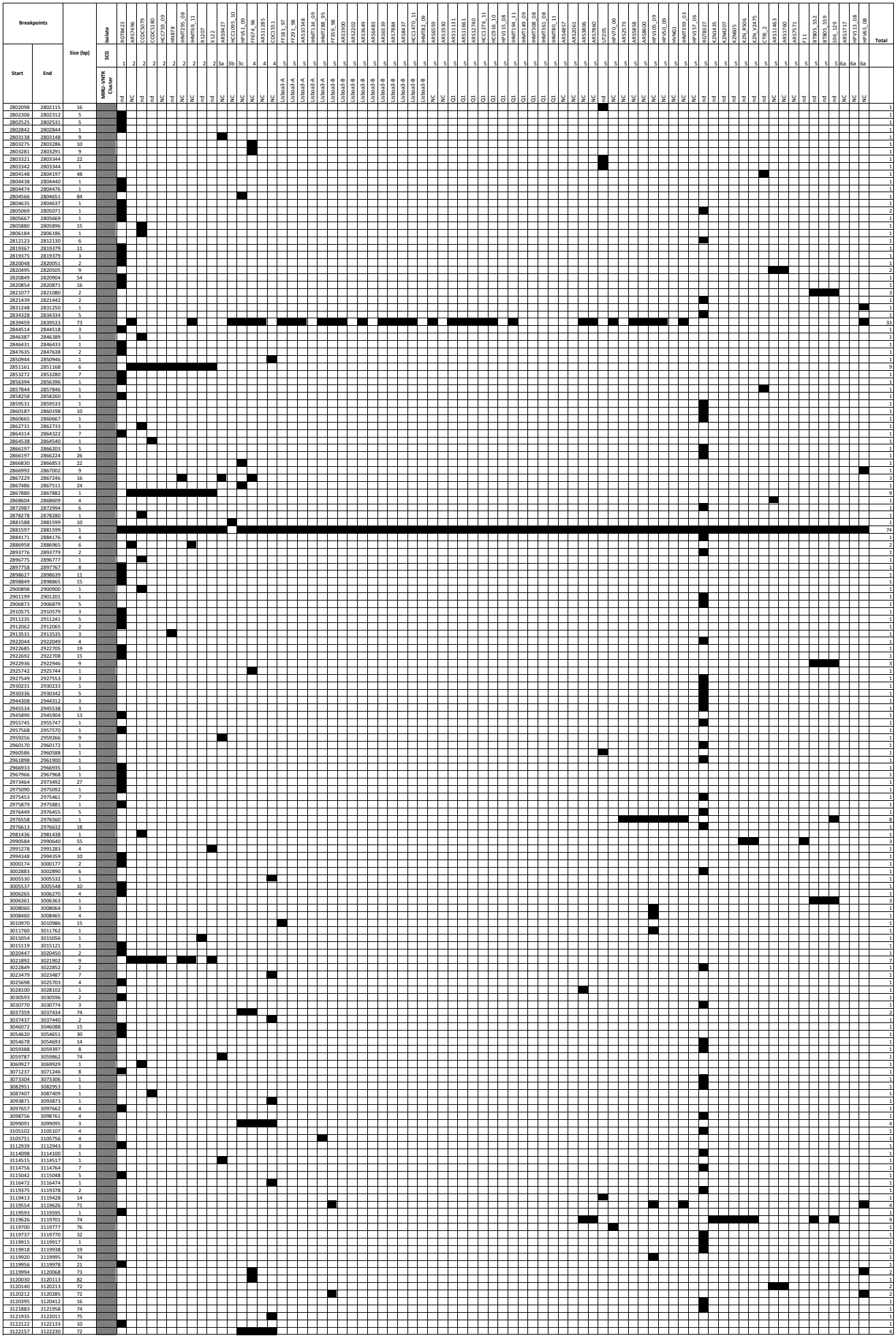

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NMNJ cNTR Cluster
Total

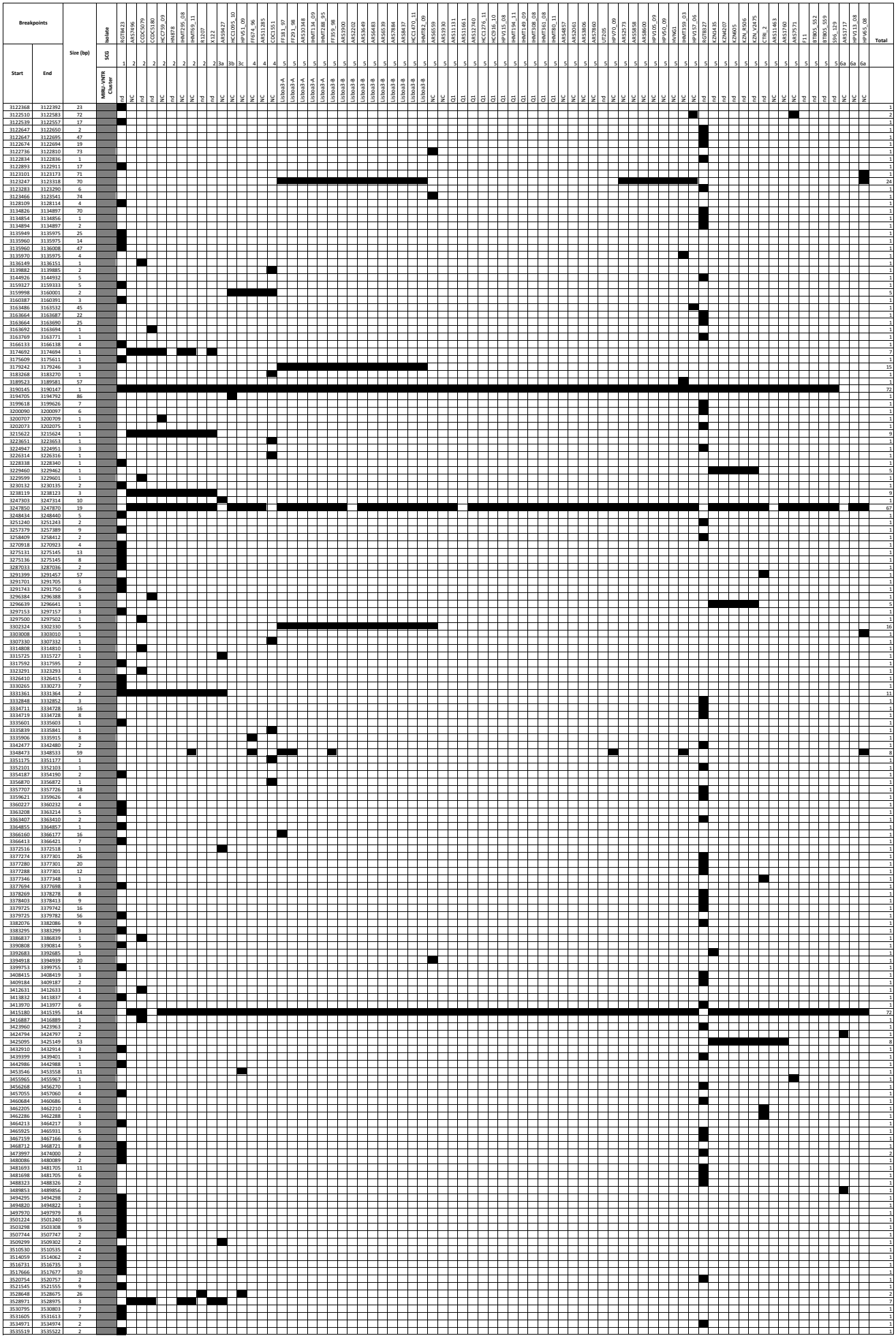

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MRU-LHVTR Cluster
Total

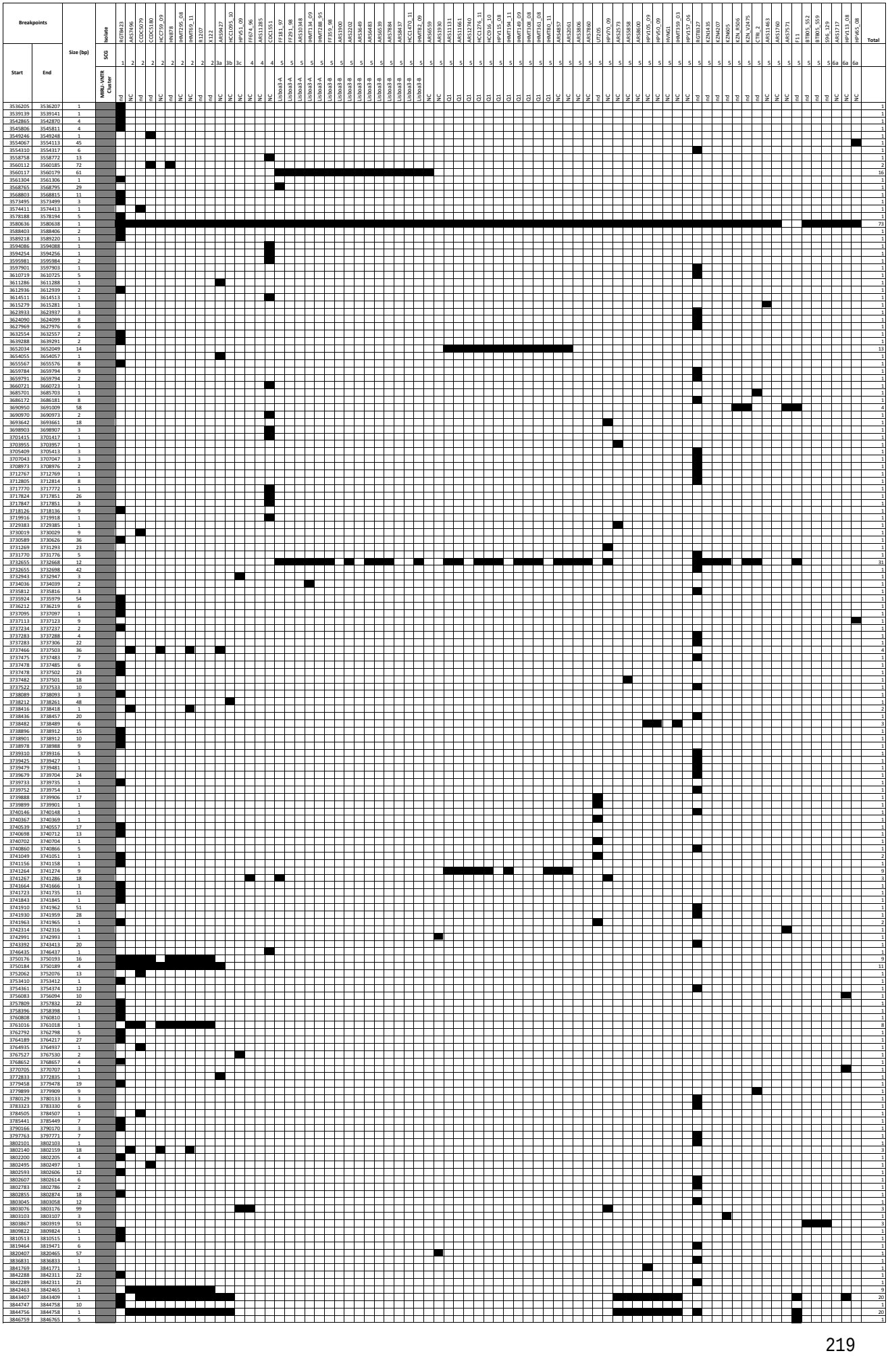

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NRMI-I/HTR Cluster
Total

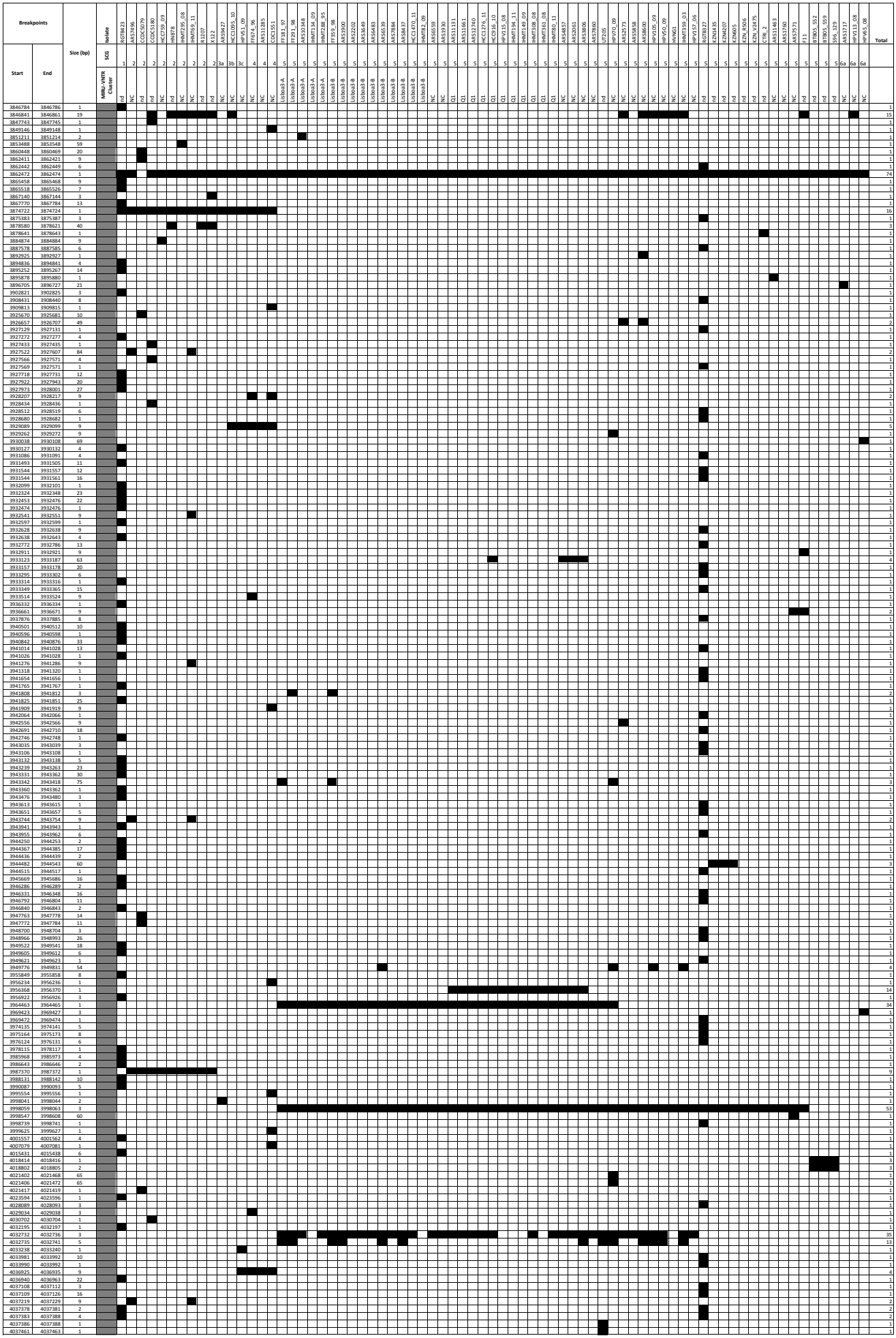

Breakpoints
Start
End
Size (bp)
Isolate
SCG
NMNJ cNTR Cluster
ISOTH431
ABS75496
CGAC5079
CGAC5180
HCC739_09
HNH78
HMT295_08
HMT69_11
R1207
X122
ABS9427
HCC1095_10
HPV51_09
FF674_96
ABS11285
CGC1551
FF181_97
FF291_98
ABS91348
HMT134_09
HMT288_05
FF393_98
ABS19300
ABS2202
ABS98149
ABS6483
ABS65139
ABS7884
ABS8437
HCC1470_11
HMT82_09
ABS6559
ABS1930
ABS11131
ABS11661
ABS12740
HCC1276_11
HCC936_10
HPV115_08
HMT194_11
HMT149_09
HMT108_08
HMT160_08
HMT80_11
ABS4857
ABS2061
ABS3806
ABS7860
UTZ05
HPV70_09
ABS2573
ABS3858
ABS8600
HPV105_09
HPV50_09
HIV061
HMT139_03
HPV157_06
ISOTH077
KZN2435
KZN4207
KZN605
KZN_K906
KZN_V2475
CTR_2
ABS11463
ABS3760
ABS791
F11
BTB05_552
BTB05_559
596_129
ABS1717
HPV113_08
HPV65_08
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MRM-LHVFR Cluster
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MRM-I-VNTR Cluster
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MRU-LHVR Cluster
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MRU-LHVFR Cluster
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
NRMI-I/HTR Cluster
Total
1
2
3a
3b
3c
4
5
6a
6b
nd
NC
Q1

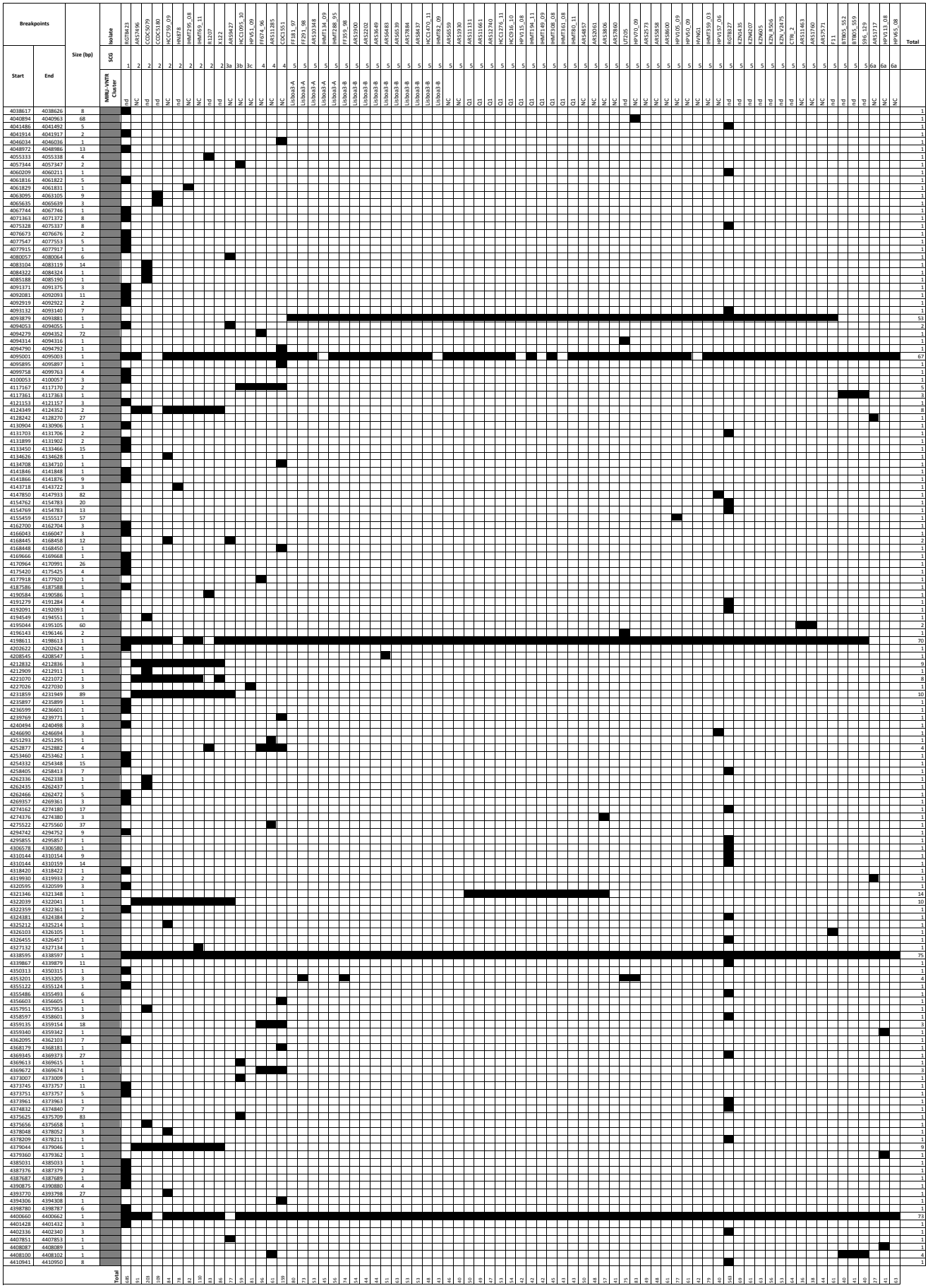
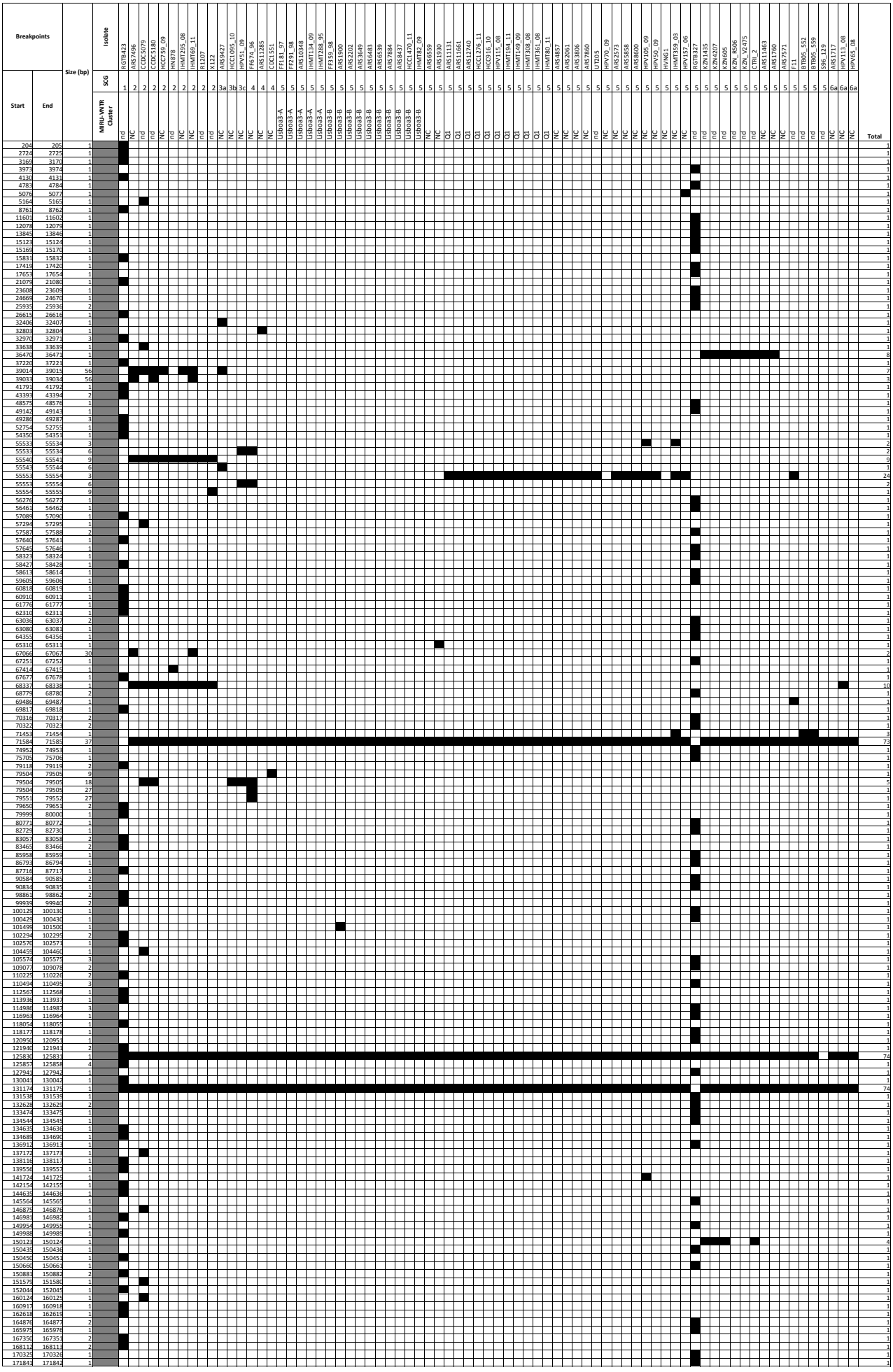
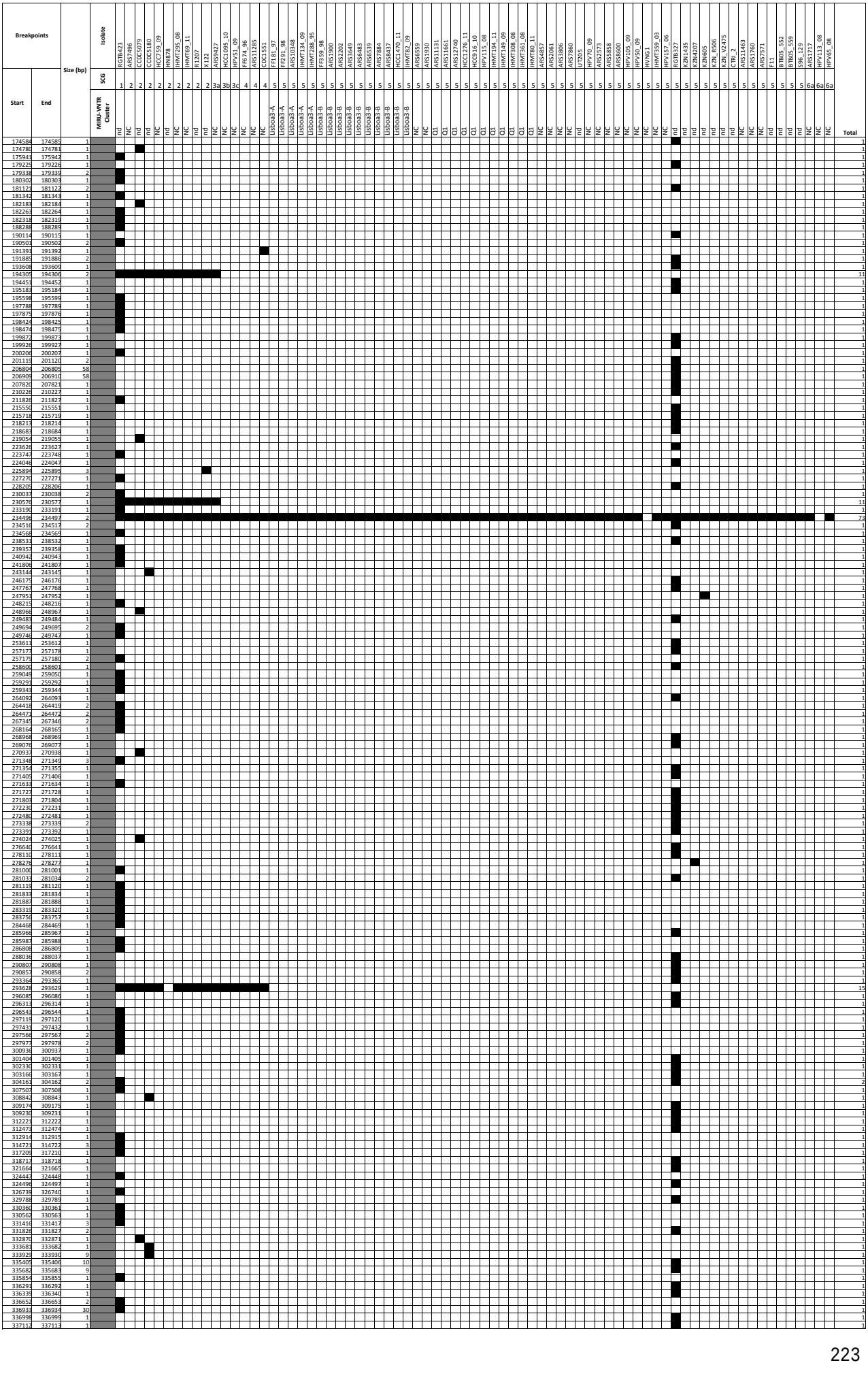
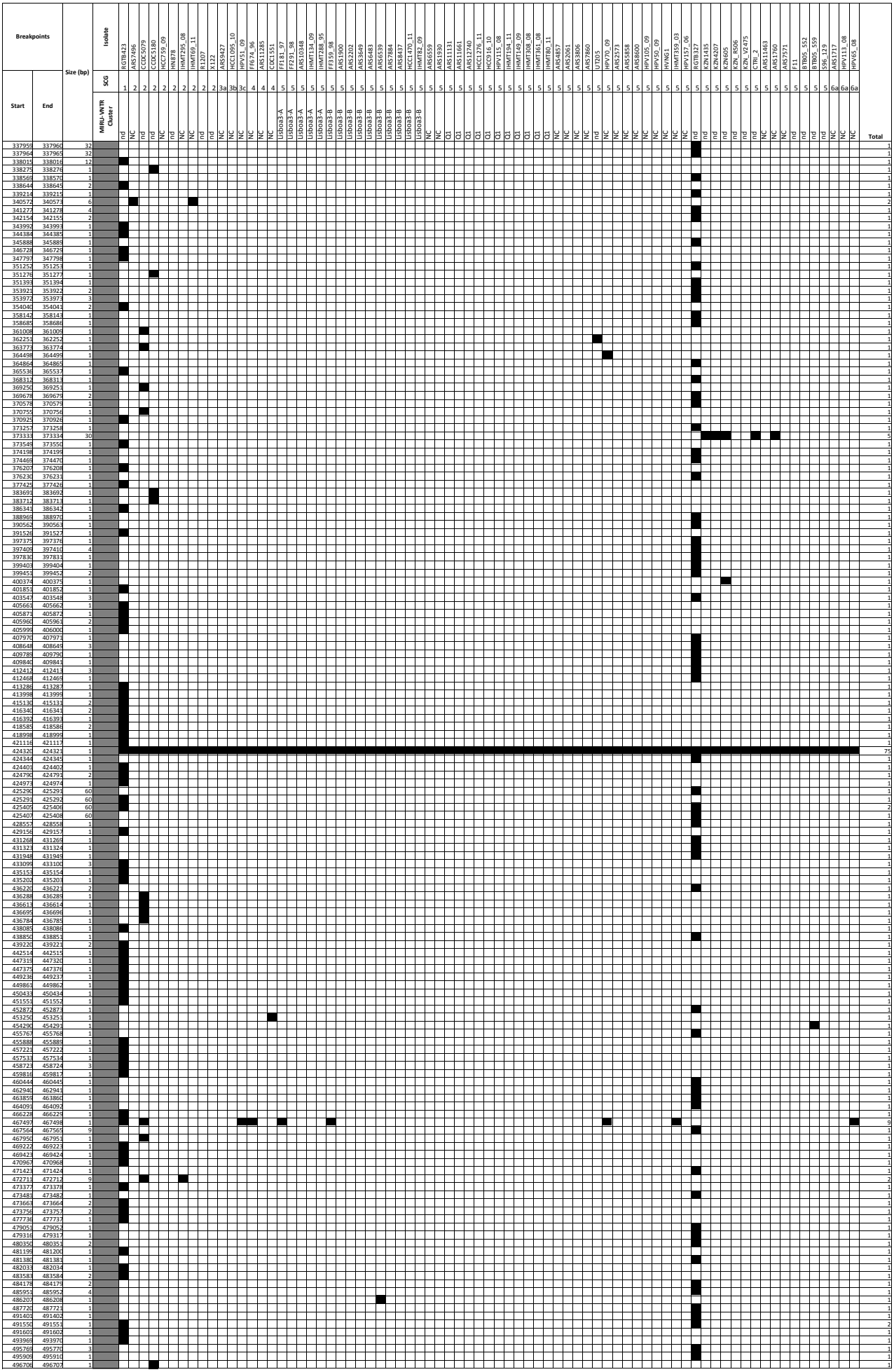
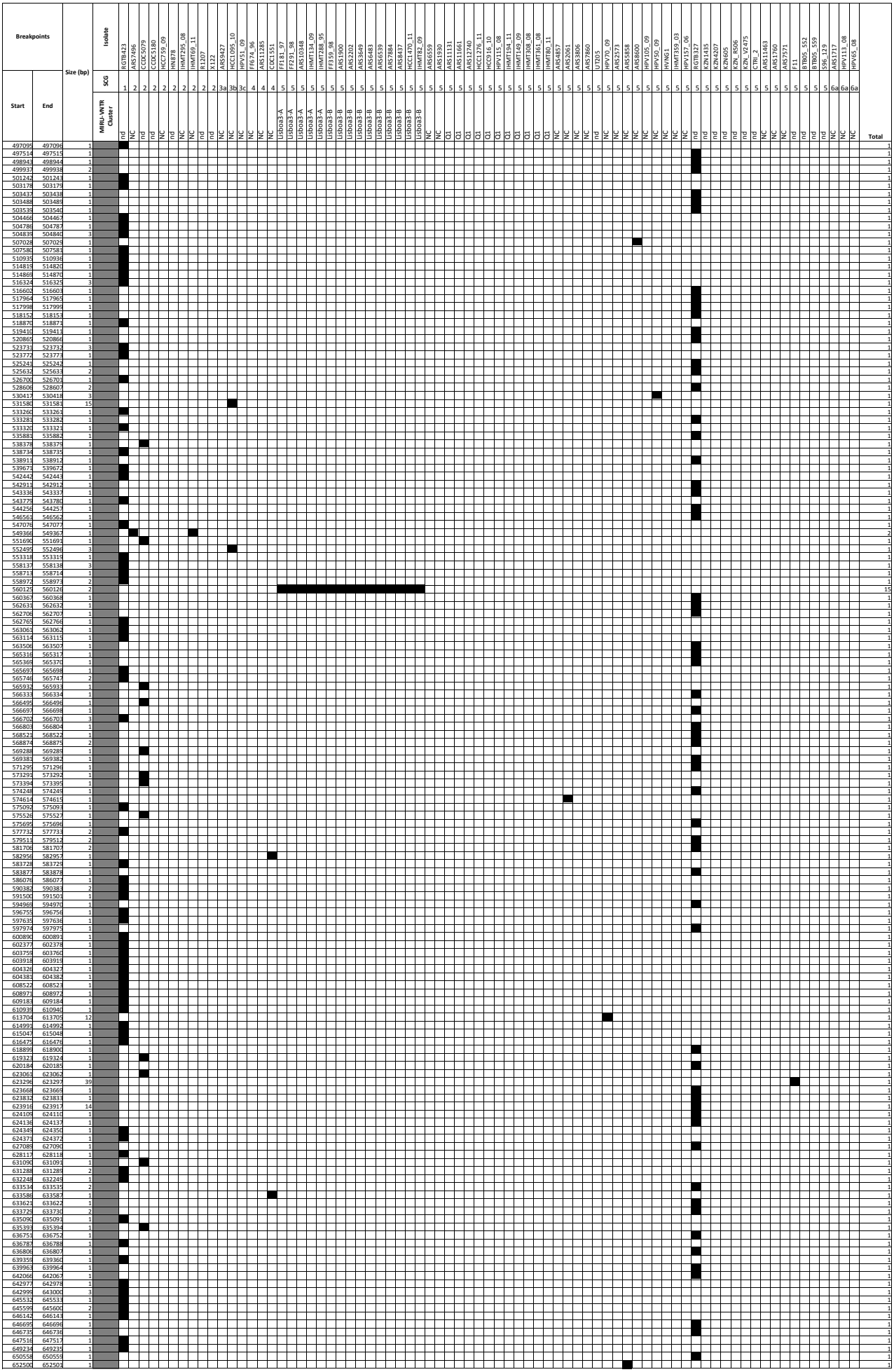


Figure S7.5 - List of short deletions (<100 bp) found among the group of 75 clinical isolates. Black squares indicate deletion detection. MIRU-VNTR cluster indicates the 24-loci MIRU-VNTR cluster of any given isolate, except if non-clustered (NC) or not determined (nd). Line and column totals indicate total column/line count of deletions.

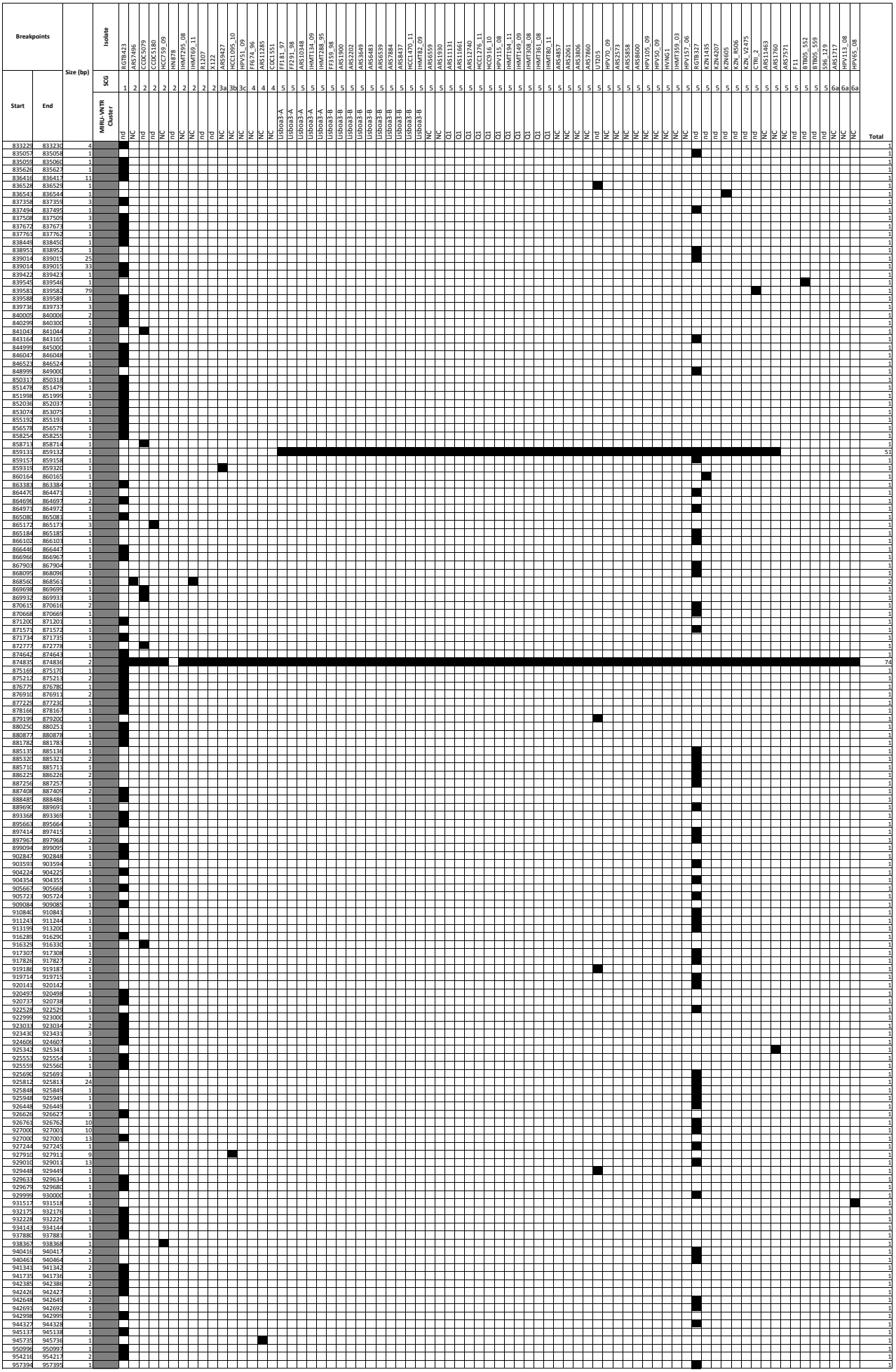

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
FF191_07
FF191_98
ARS10348
HMIT134_09
HMIT288_95
FF159_98
ARS19900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVIN61
IHMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

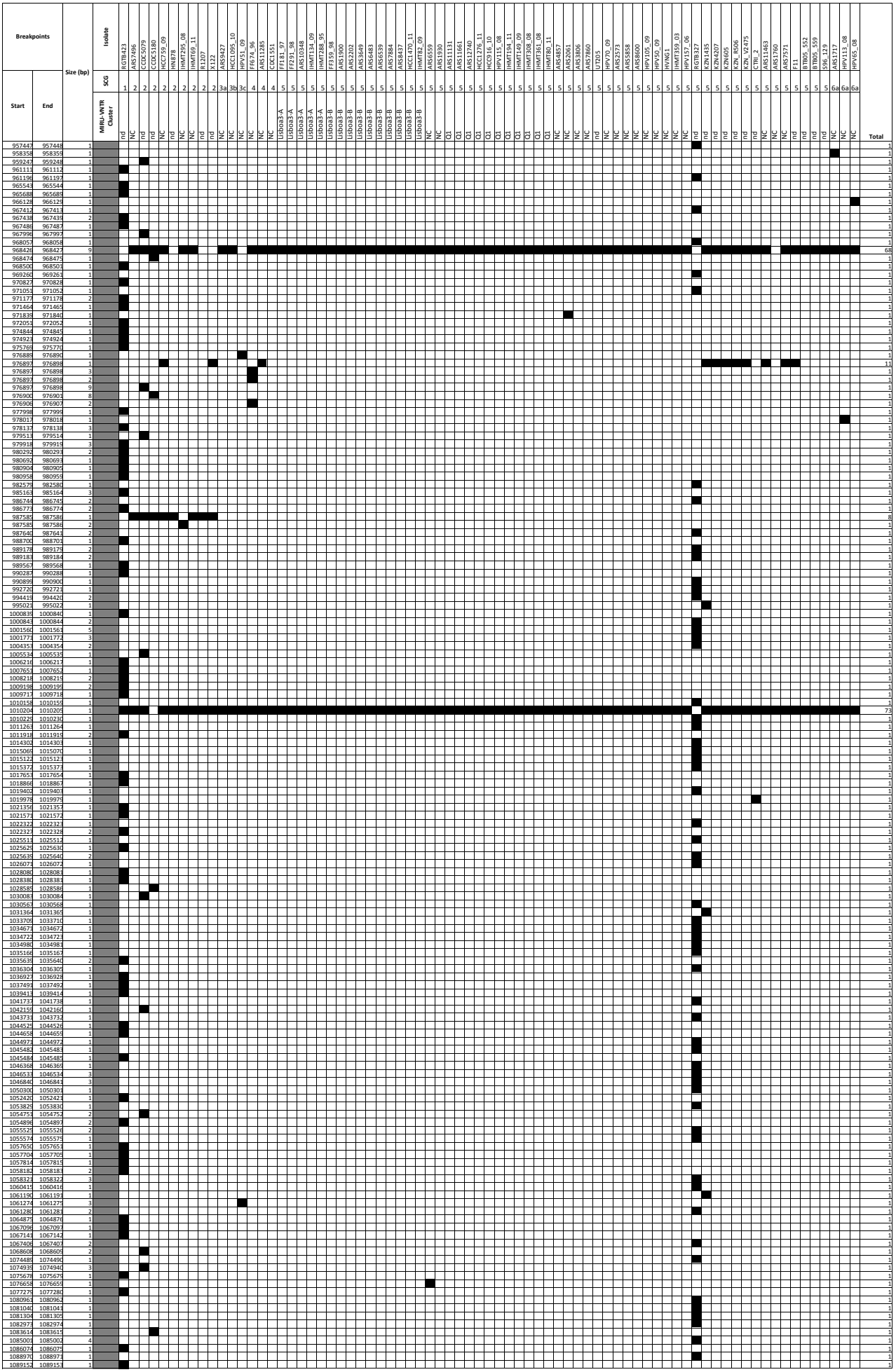

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11635
CDC1551
FJ191_97
FF191_98
ARS10348
HMIT194_09
HMIT288_95
FF159_98
ARS19000
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HVIN61
HMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

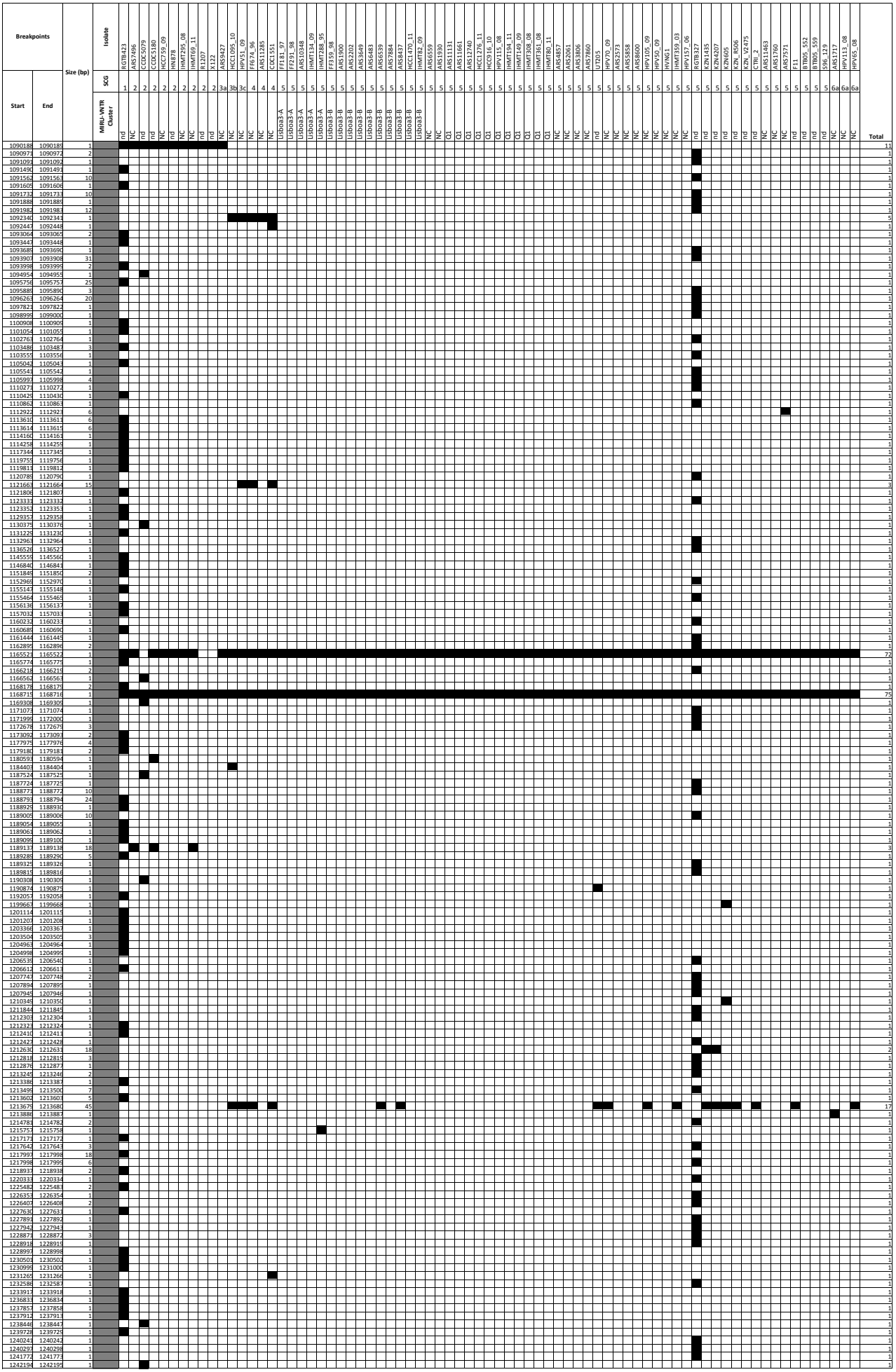

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV517_09
FF674_96
ARS11635
CDC1551
FF191_97
FF191_98
ARS10348
HMT134_09
HMT288_95
FF159_98
ARS19900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRL_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

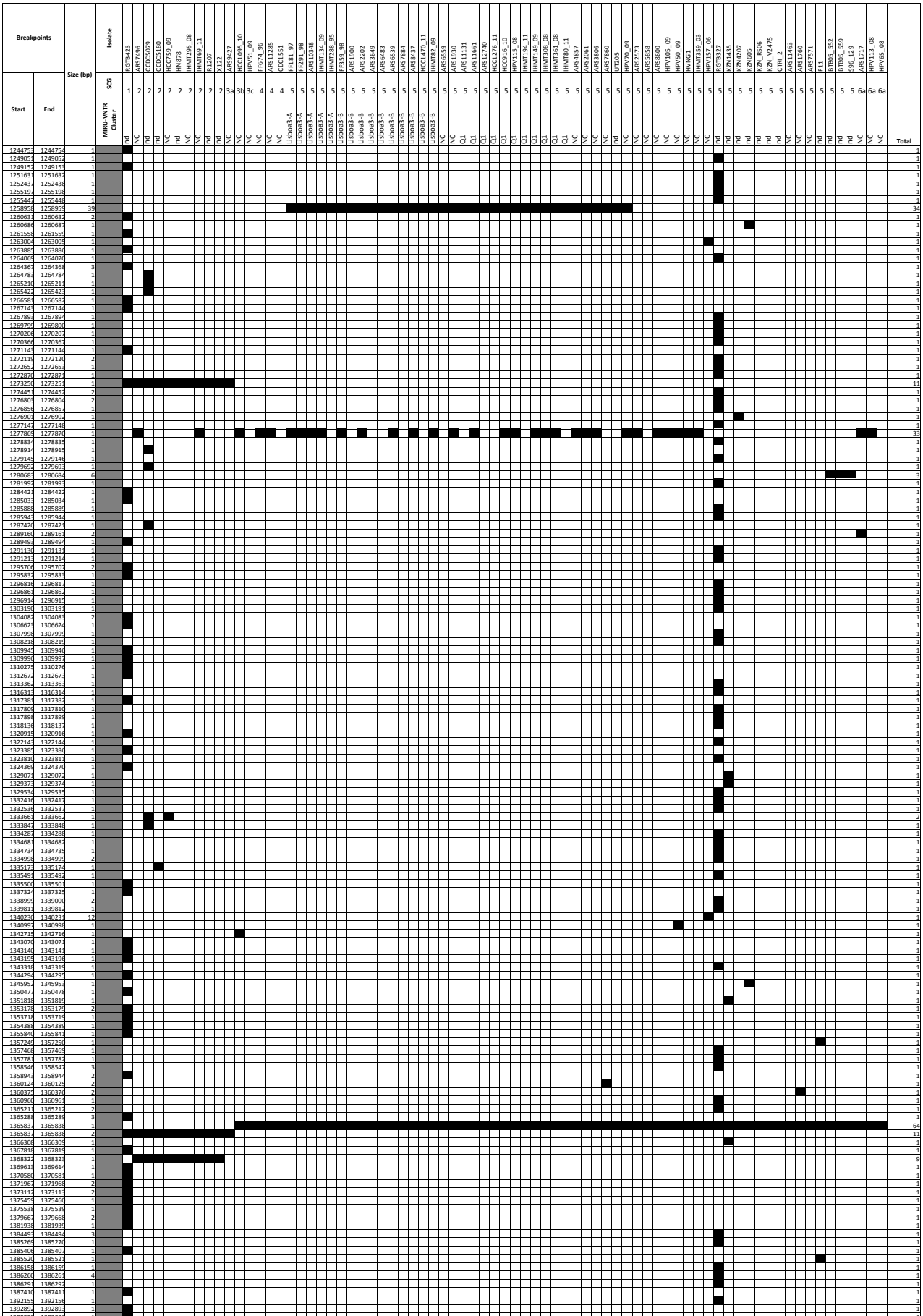

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_06
ARS11635
CDC1551
FF191_07
FF191_08
ARS10348
HMT194_09
HMT288_05
FF159_08
ARS19000
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HIVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRL_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

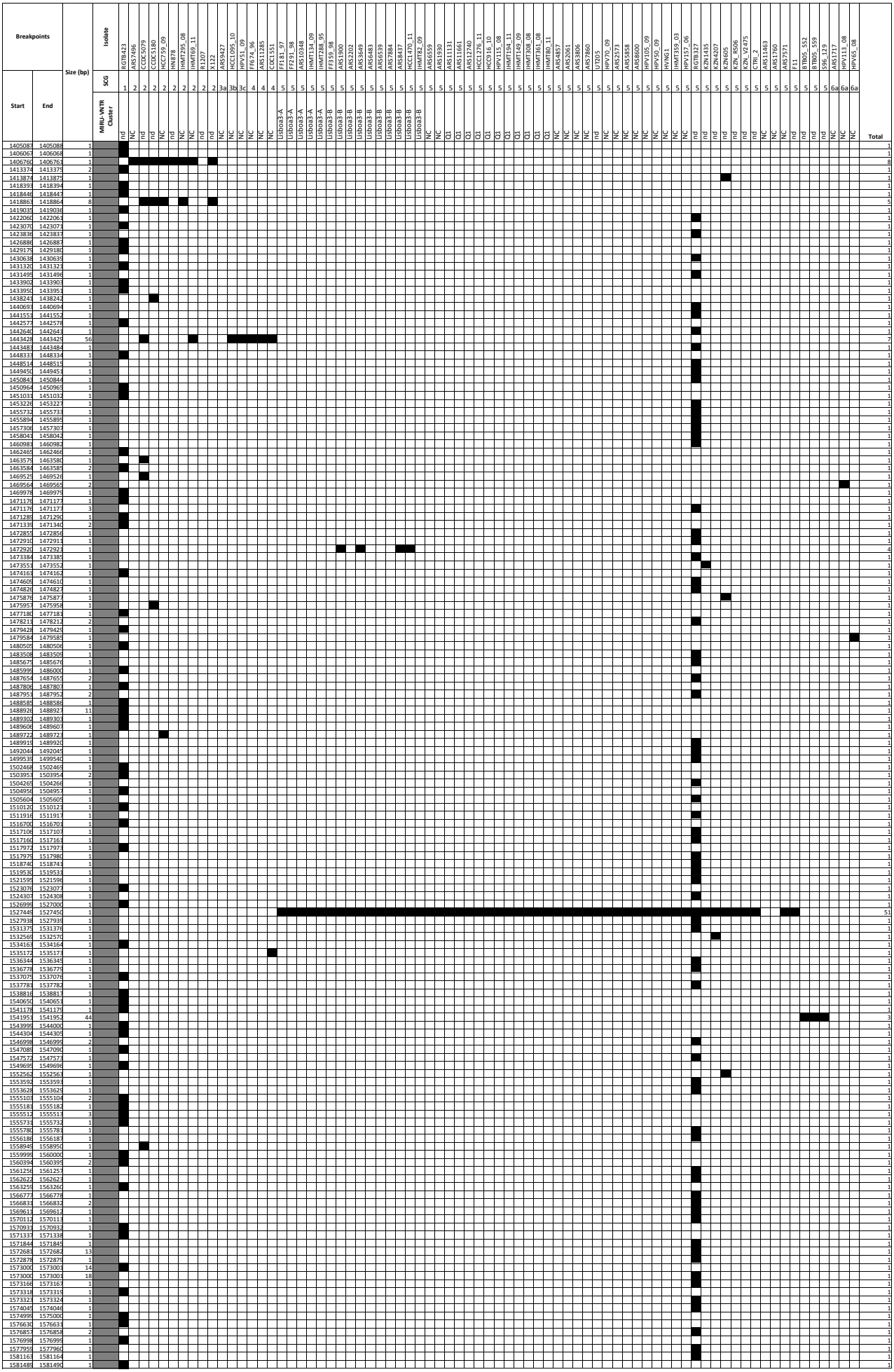

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
FF191_07
FF191_98
ARS10348
HMIT134_09
HMIT288_95
FF159_98
ARS19900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

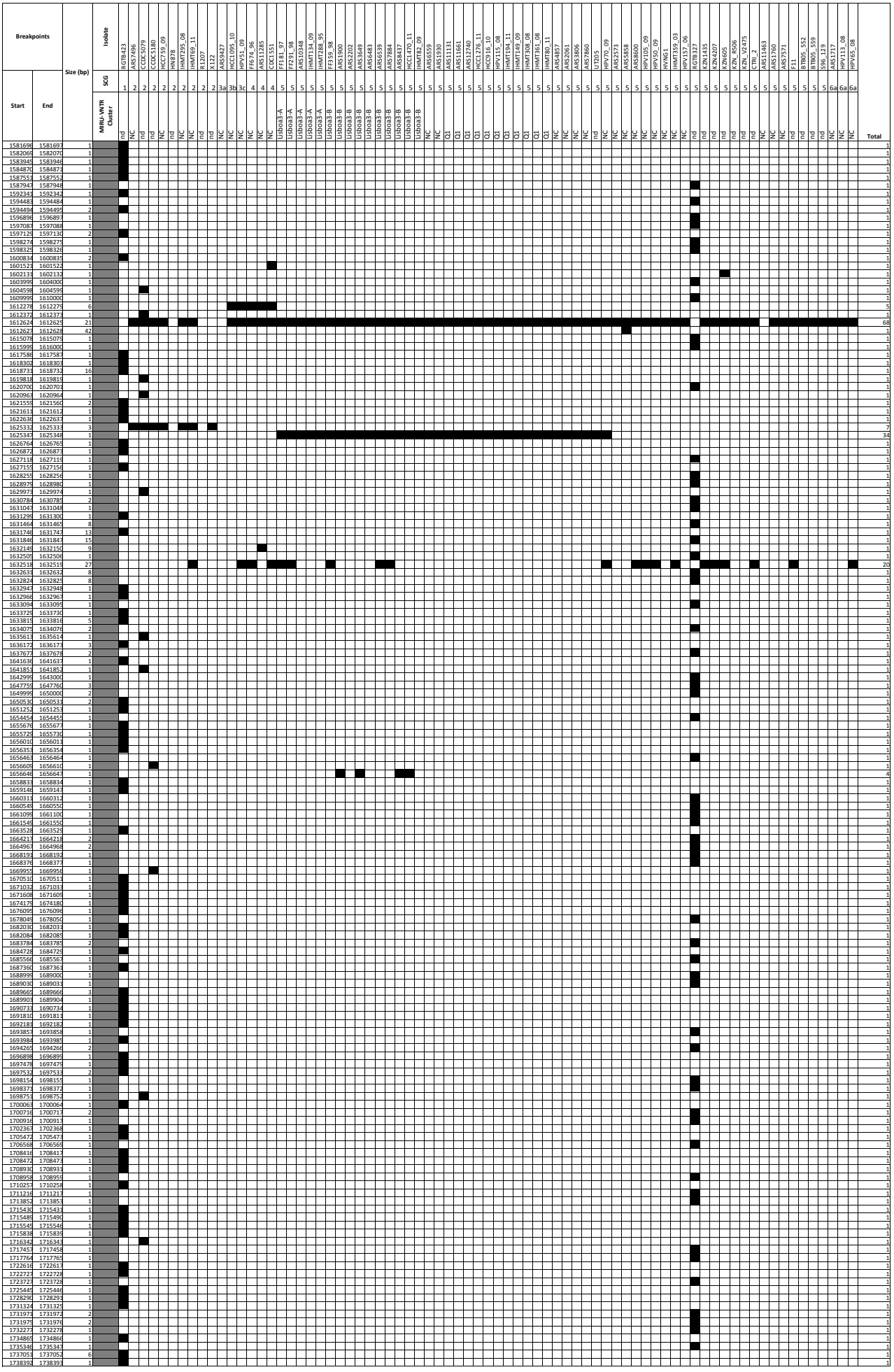

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11635
CDC1551
U1dos3-A
U1dos3-A
U1dos3-A
U1dos3-A
ARS10348
HMT134_09
HMT288_95
FF159_98
ARS15900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HIVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRL_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

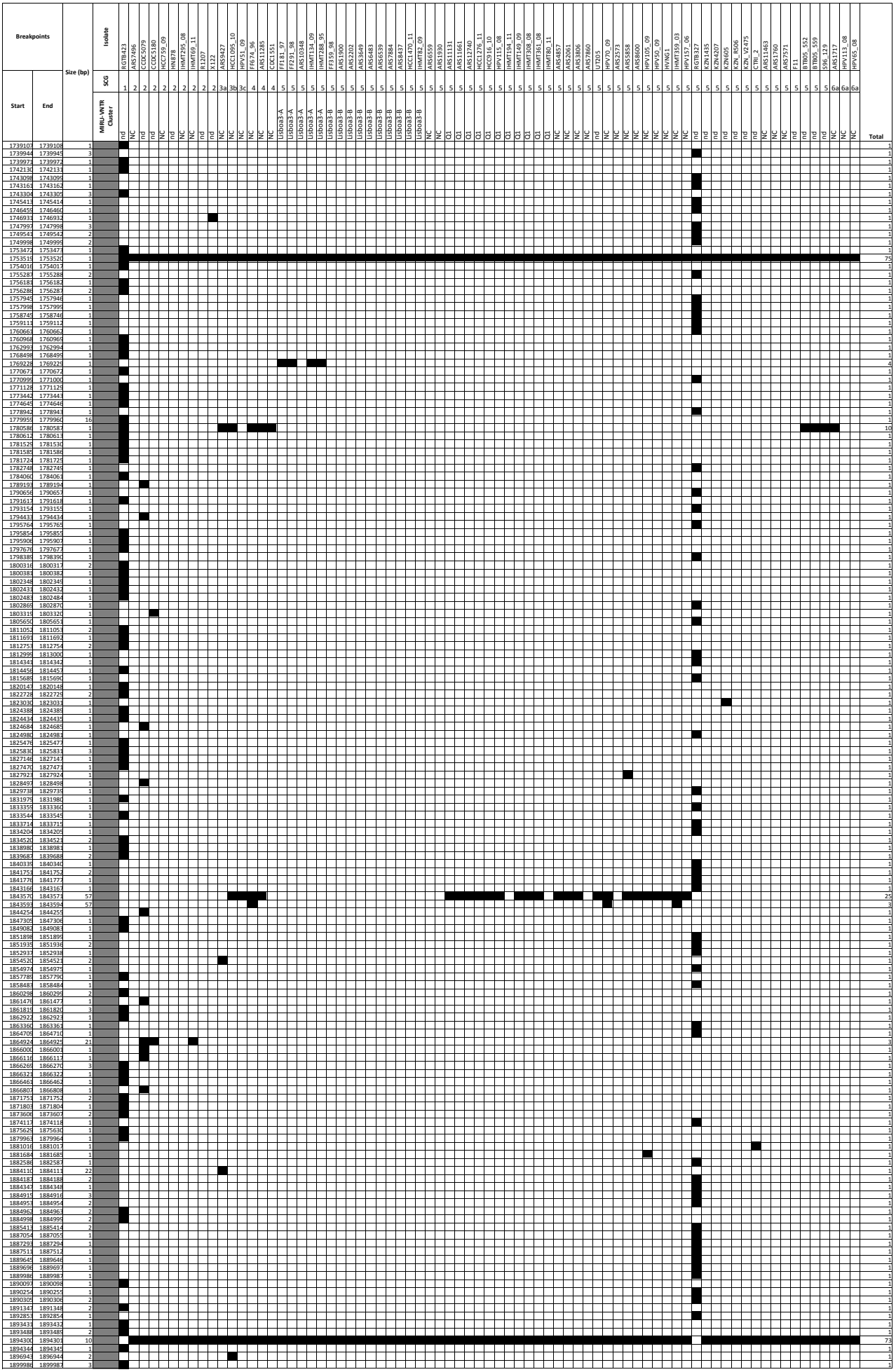

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HNB78
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_06
ARS11835
CDC1551
FF191_07
FF191_08
ARS10348
HMT194_09
HMT288_05
FF159_08
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

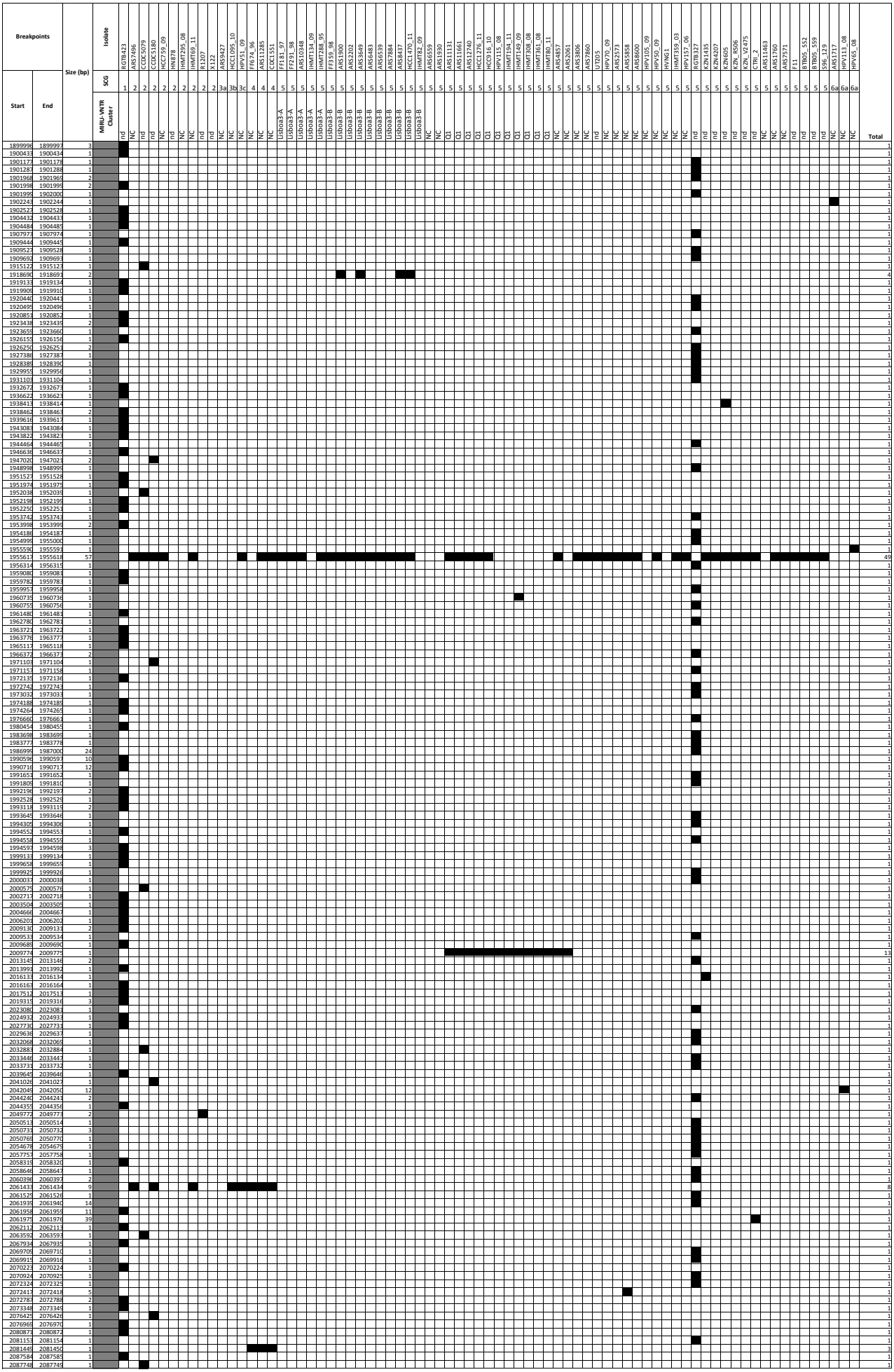

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
AR59427
HCC1095_10
HPV157_09
FF674_06
AR511835
CDC1551
FT191_07
FT191_08
AR510348
HMT134_09
HMT288_05
FT159_08
AR512900
AR52202
AR53849
AR56483
AR56539
AR57884
AR58437
HCC1470_11
HMT82_09
AR56559
AR51930
AR511131
AR511661
AR512740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
AR54857
AR52061
AR53806
AR57860
UT205
HPV70_09
AR52573
AR55858
AR58600
HPV105_09
HPV90_09
HVIN61
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
AR511463
AR51760
AR57571
F11
BTB05_552
BTB05_559
S96_129
Total

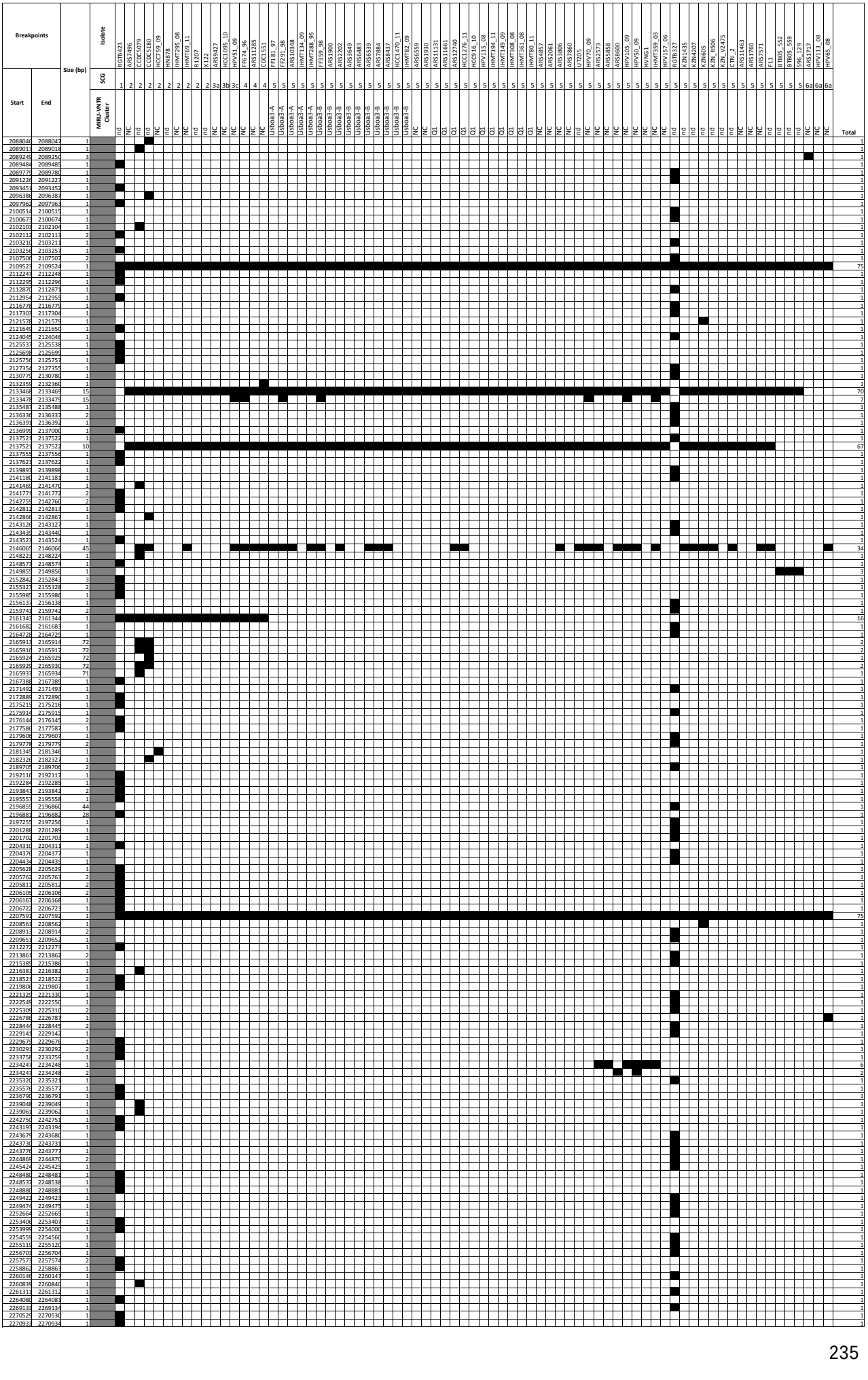

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HN878
HMT395_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
CDC1551
FF191_07
FF191_98
ARS10348
HMT134_09
HMT288_95
FF159_98
ARS19900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

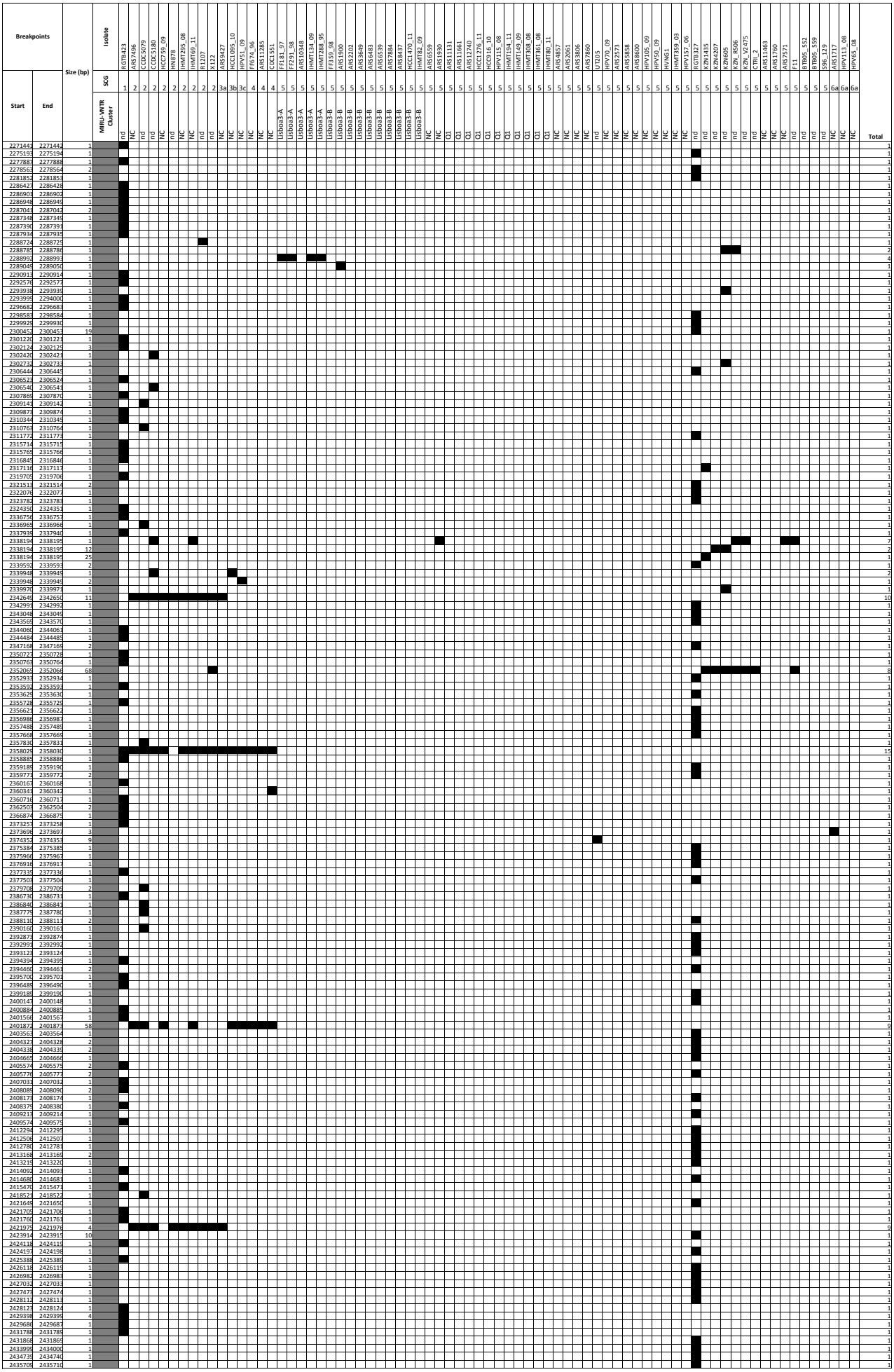

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT395_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
FT191_98
ARS10348
HMT134_09
HMT288_95
FT159_98
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HIVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

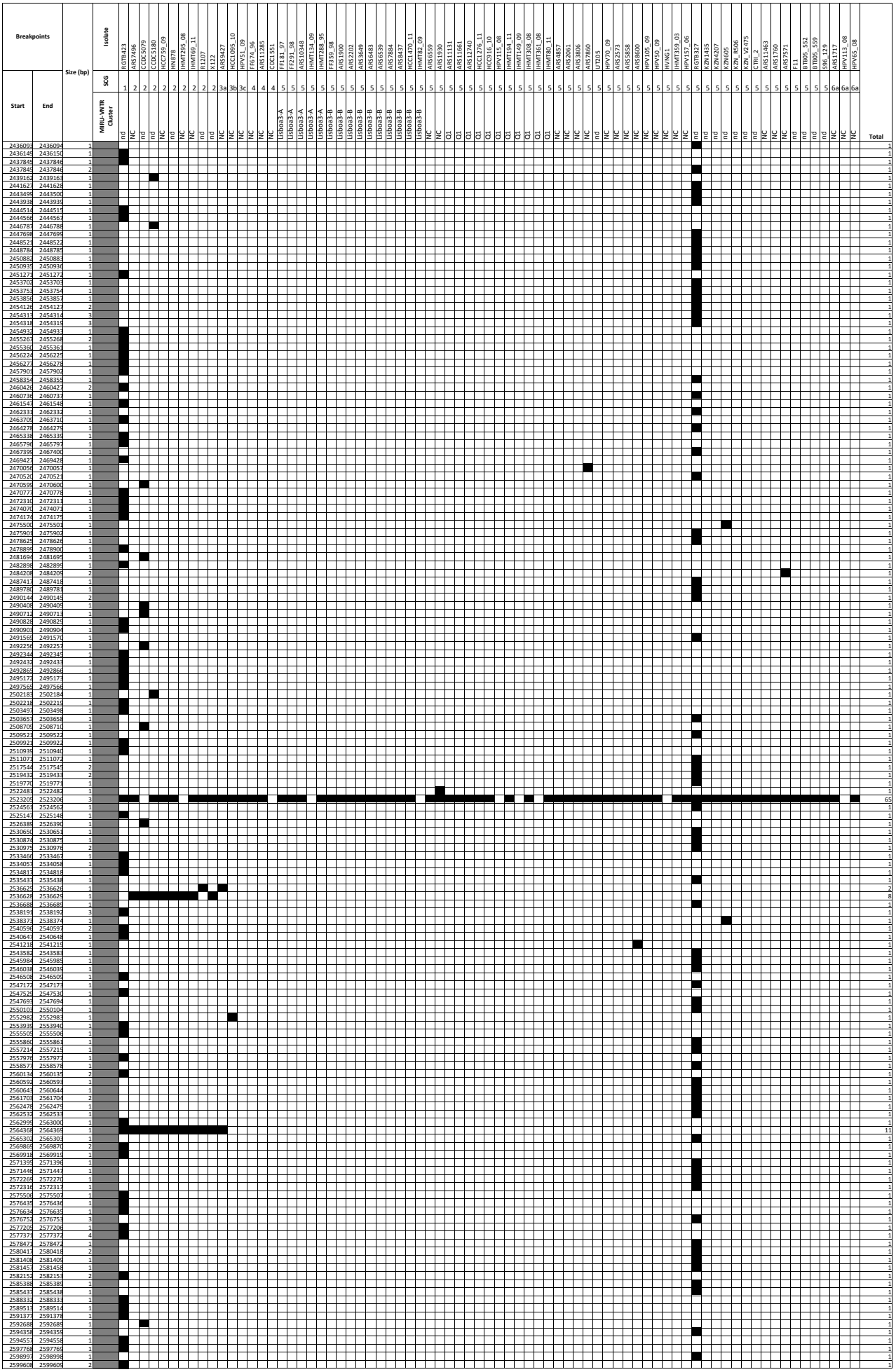

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_06
ARS11835
CDC1551
FF191_07
FF191_08
ARS10348
HMIT194_09
HMIT288_05
FF159_08
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVIN61
HMIT359_03
HPV157_06
RGTB327
K2N1435
K2N4207
K2N605
K2N_R506
K2N_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

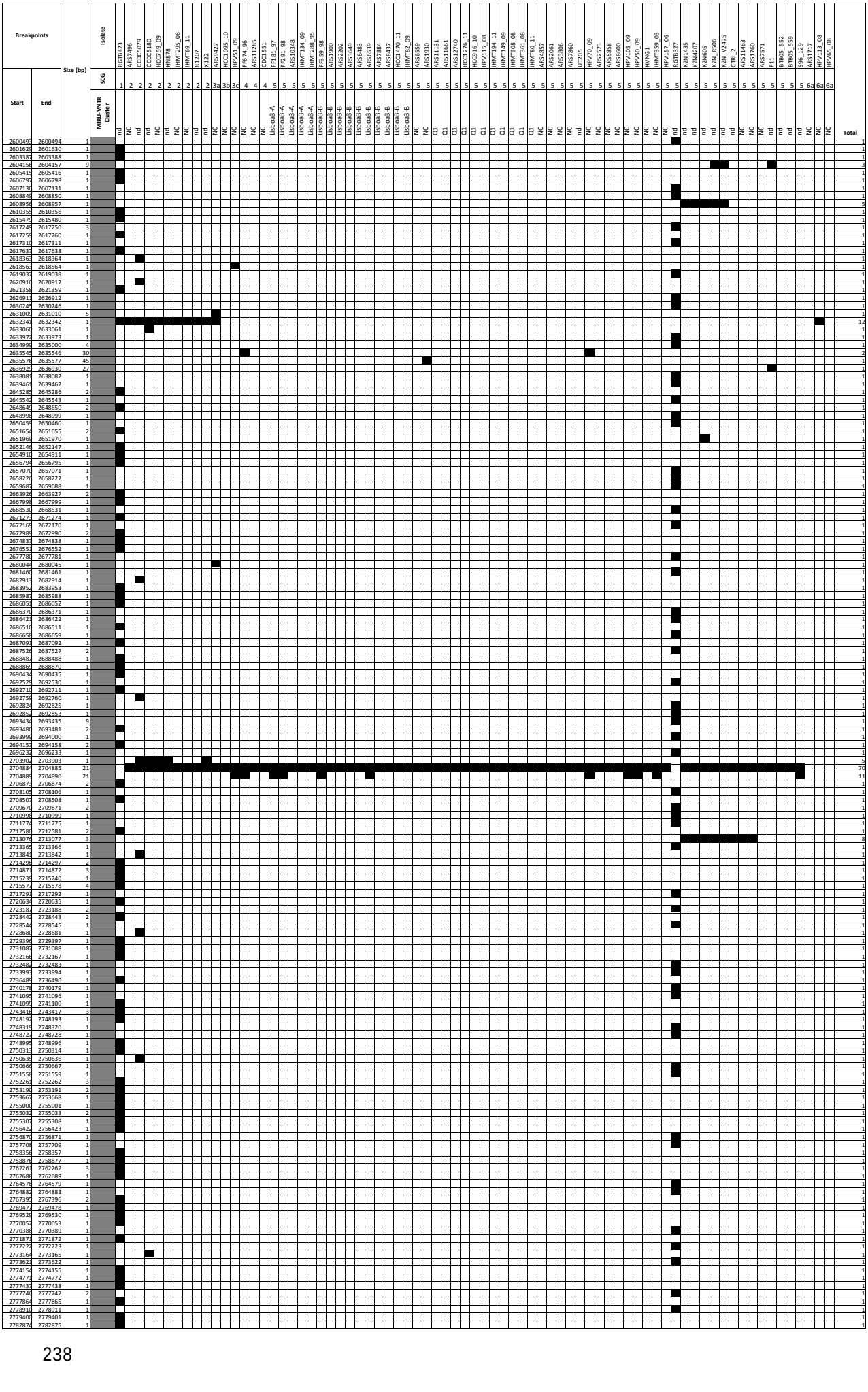

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT395_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV517_09
FF674_96
ARS11835
CDC1551
FF191_97
FF193_98
ARS10348
HMT194_09
HMT288_95
FF159_98
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

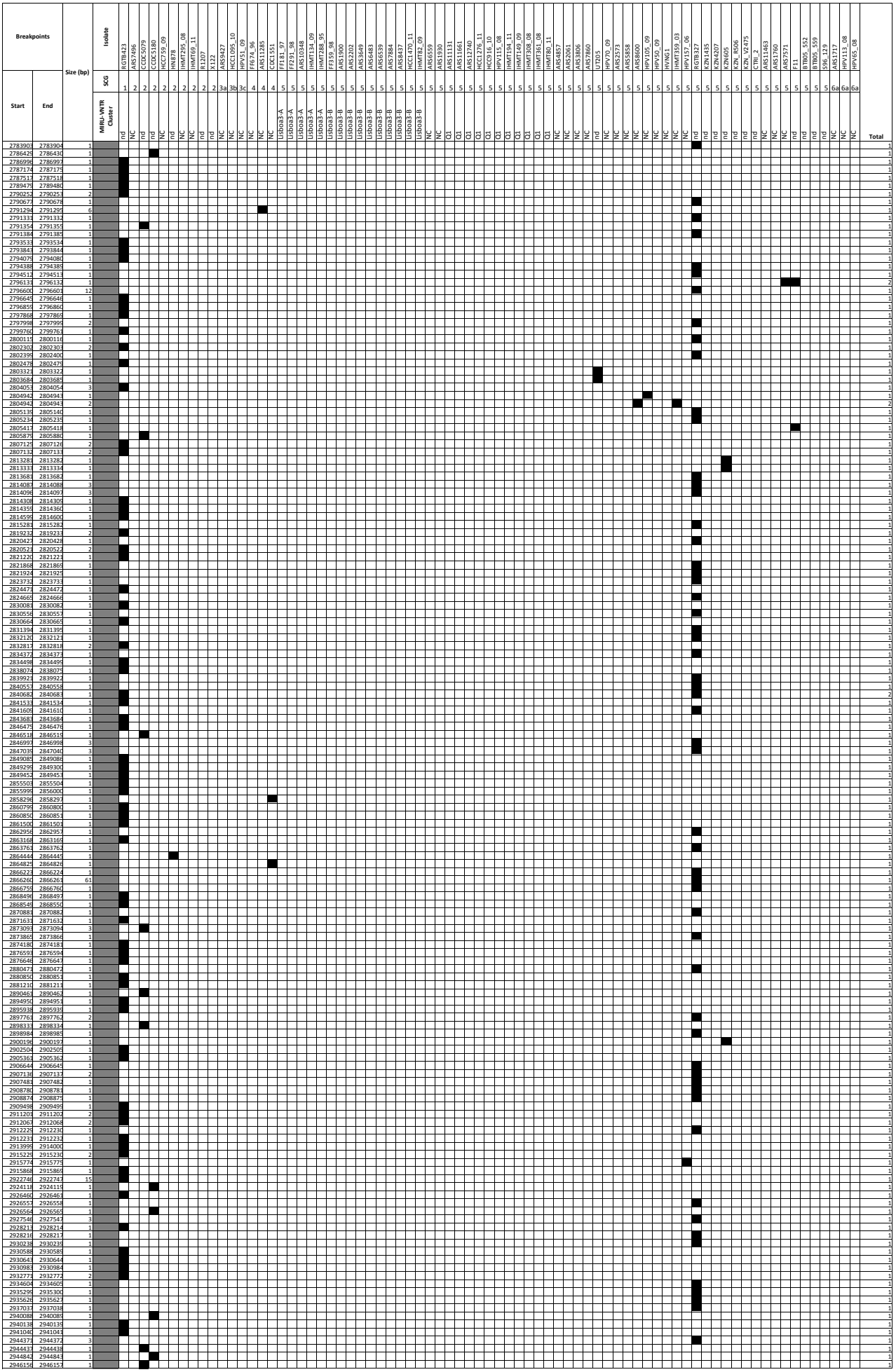

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT395_08
HMT169_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_06
ARS11635
CDC1551
FF191_07
FF191_08
ARS10348
HMT194_09
HMT288_05
FF159_08
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total

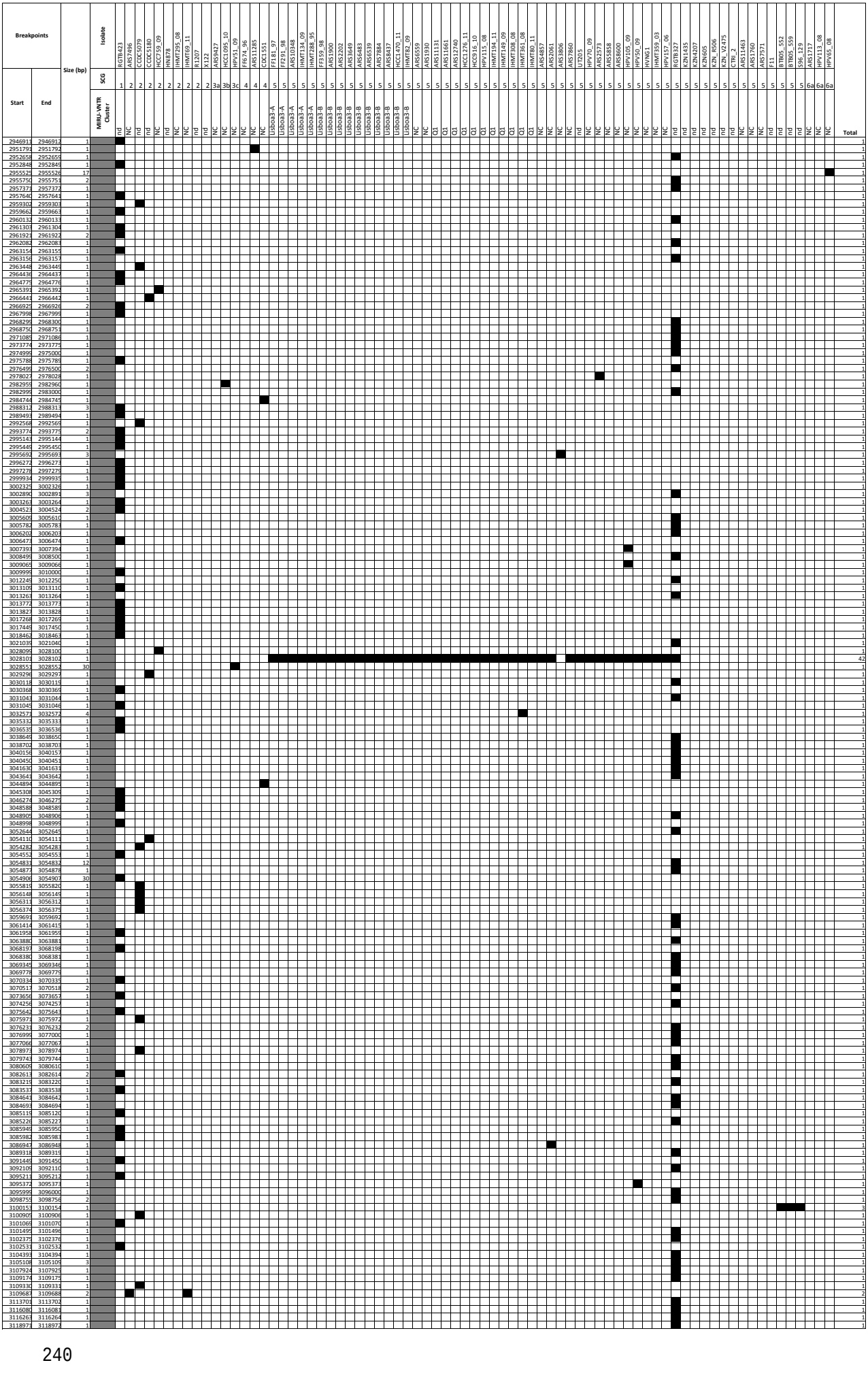

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC759_09
HN878
HMT295_08
HMT169_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
FF191_97
FF193_98
ARS10348
HMT194_09
HMT288_95
FF159_98
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRL_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
nd
NC
Total

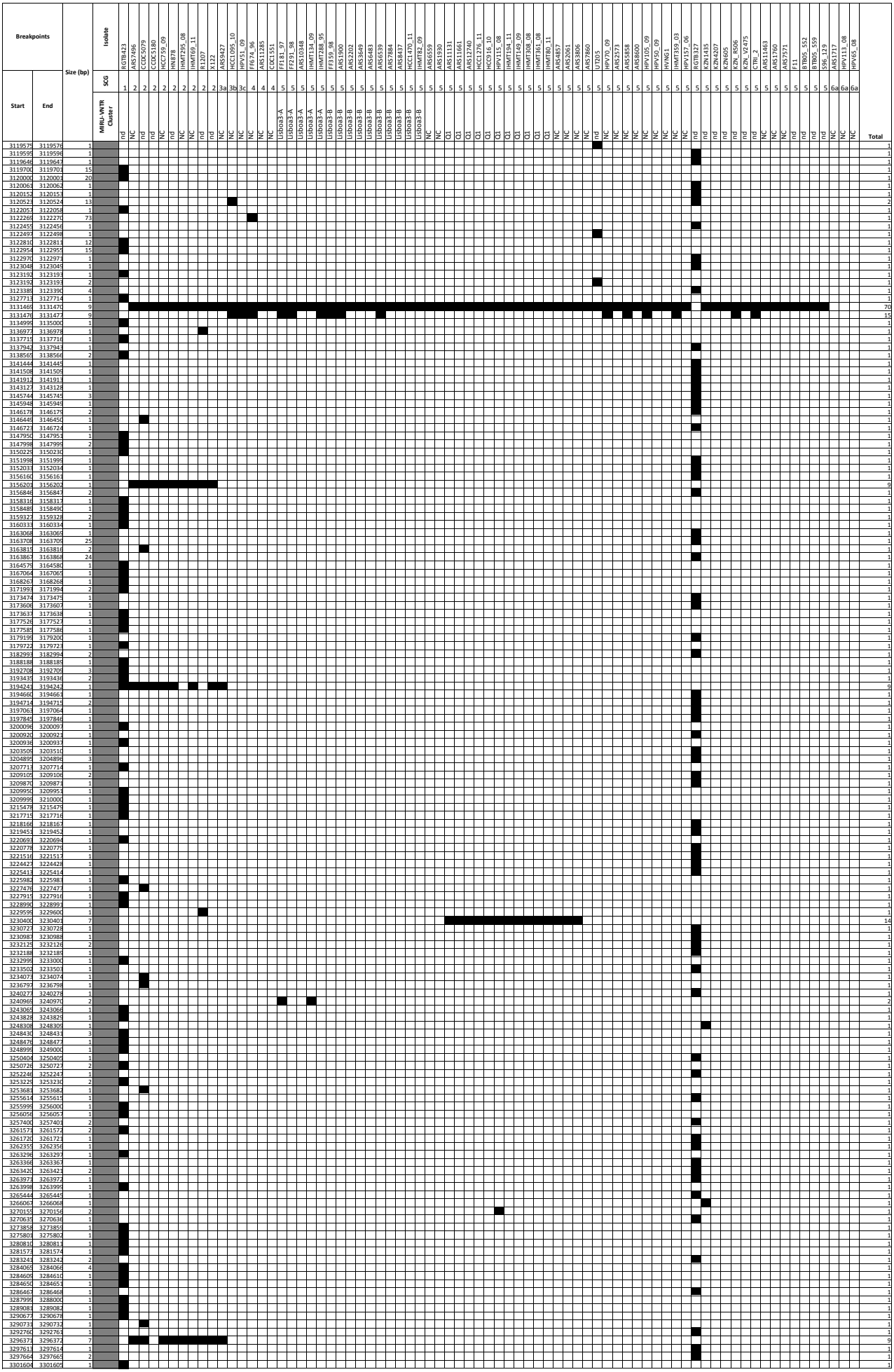

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11635
CDC1551
FF191_97
FF191_98
ARS10348
HMIT194_09
HMIT288_95
FF159_98
ARS19900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVIN61
IHMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

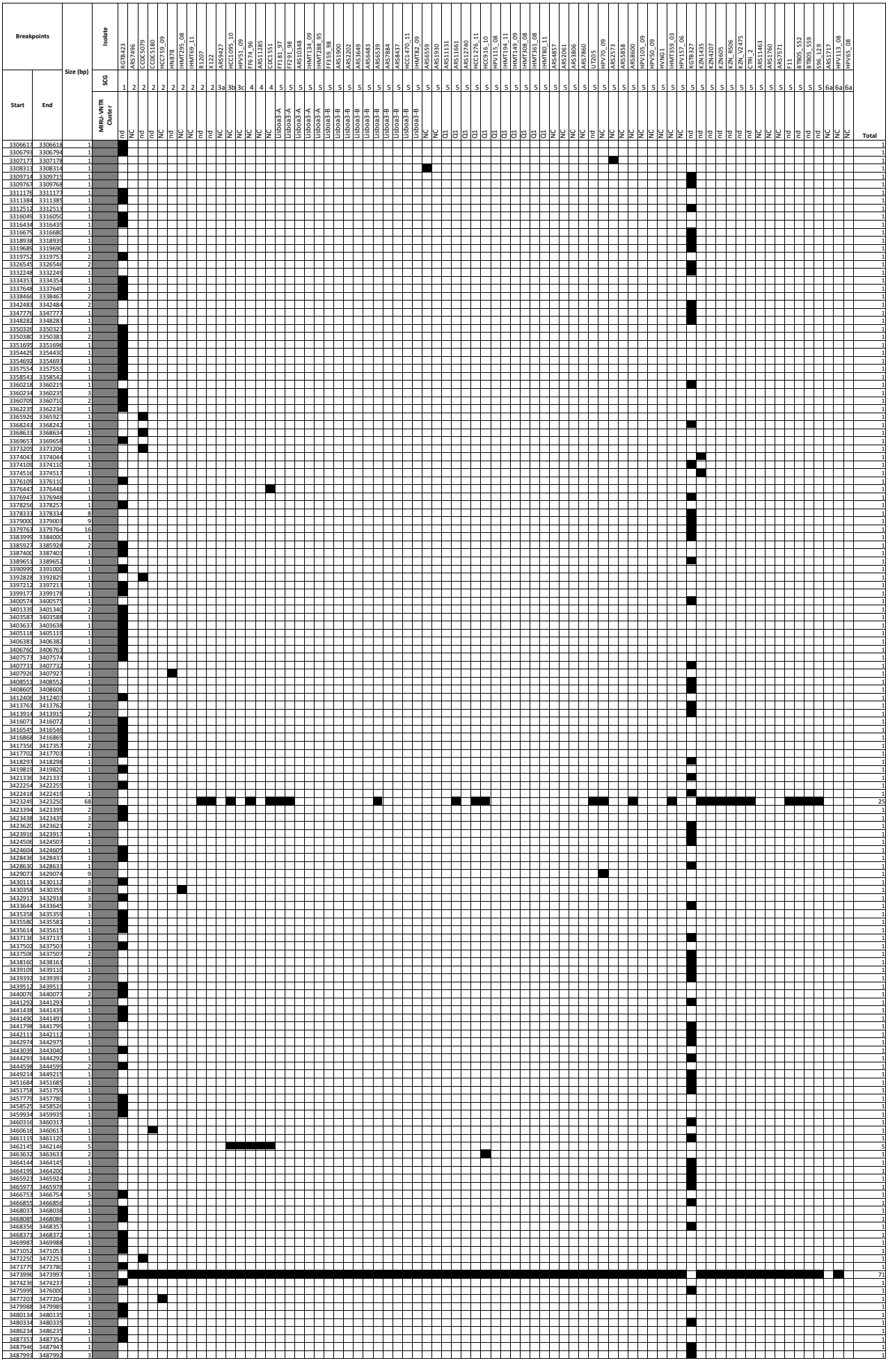

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
KCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
CDC1551
FF191_97
FF191_98
ARS10348
HMT194_09
HMT288_95
FF159_98
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HIVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HN878
HMIT295_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
HMIT3_97
FF191_98
ARS10348
HMIT134_09
HMIT288_95
FF159_98
ARS15900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVIN51
HMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

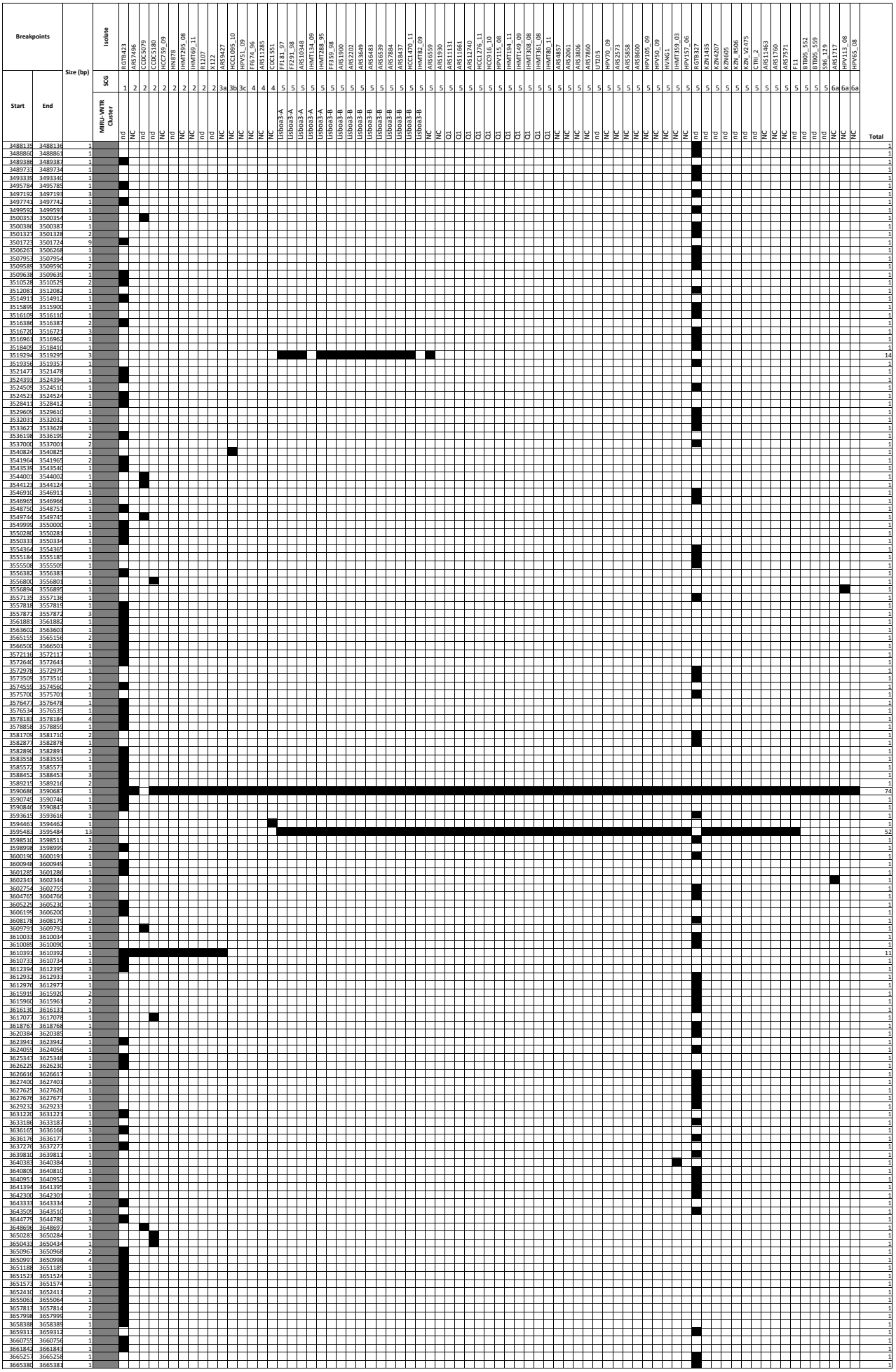

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
CDC1551
FF191_07
FF191_98
ARS10348
HMT194_09
HMT288_95
FF159_98
ARS19900
ARS2202
ARS9849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCG916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total

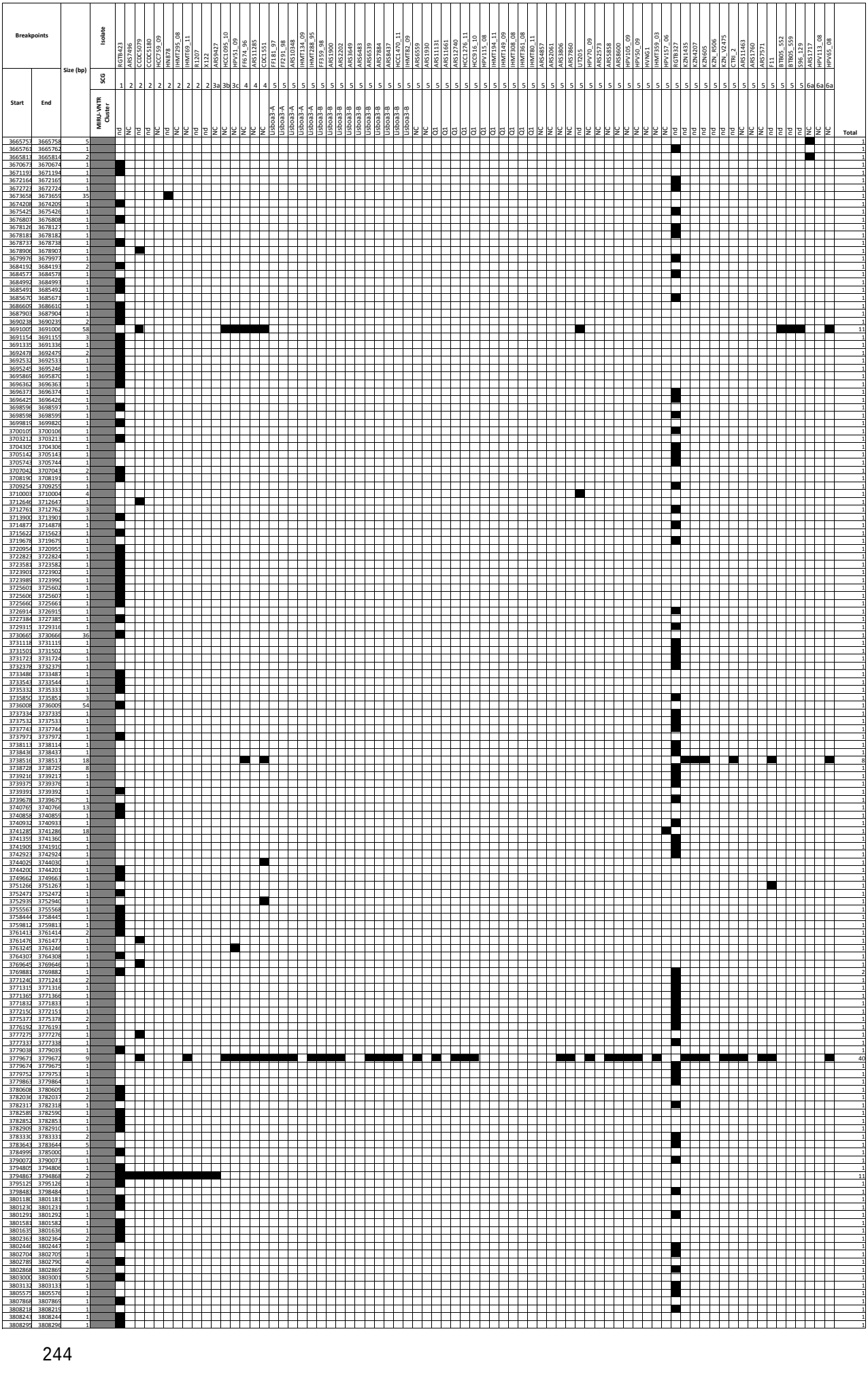

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
KCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
KCC1095_10
HPV157_09
FF674_06
ARS11835
CDC1551
FF191_07
FF191_08
ARS10348
HMT134_09
HMT288_05
FF159_08
ARS19000
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
nd
NC
Total

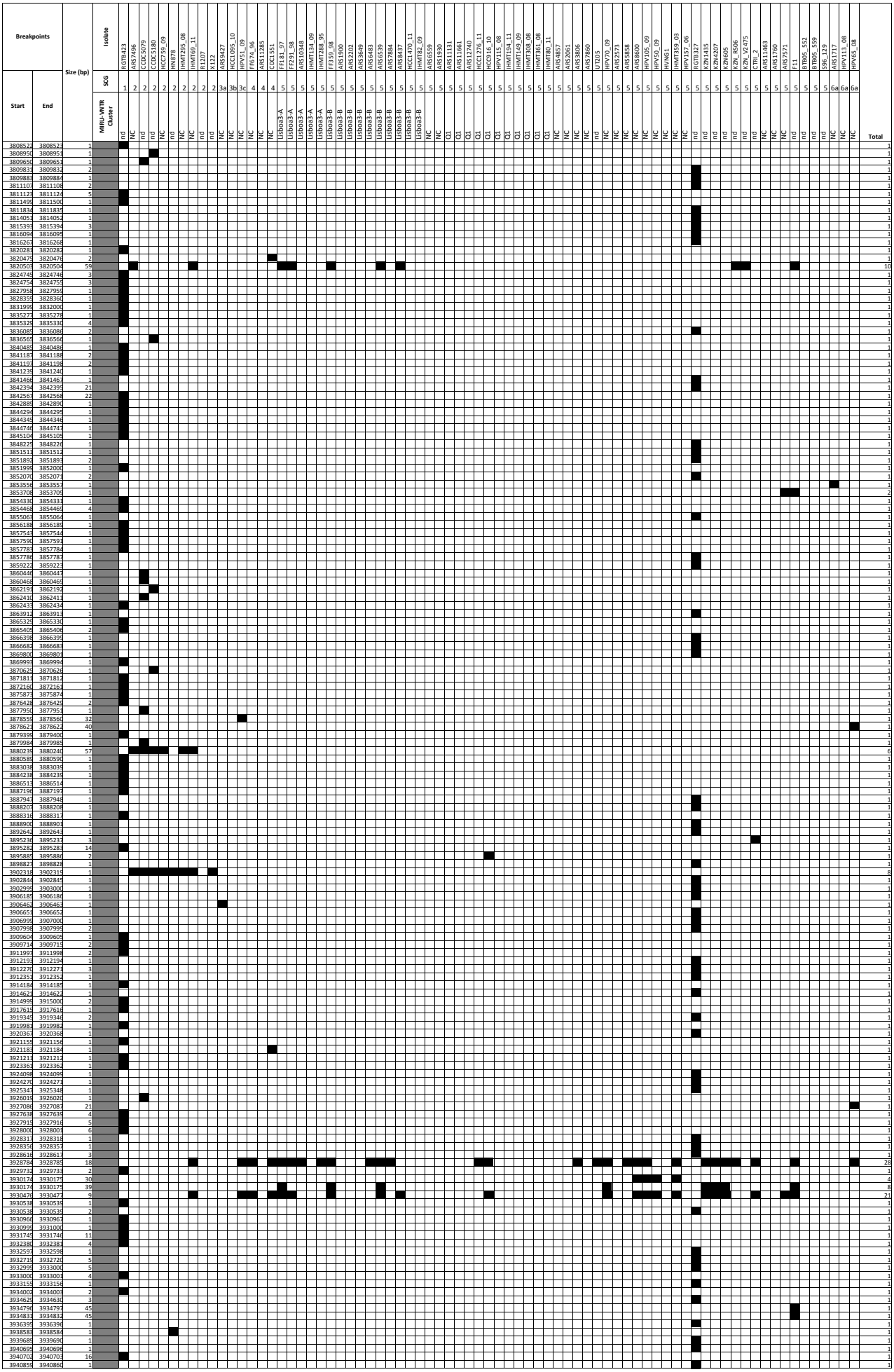

Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
CDC1551
FF191_97
FF191_98
ARS10348
HMIT194_09
HMIT288_95
FF159_98
ARS15900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV50_09
HVIN61
HMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
nd
NC
Total

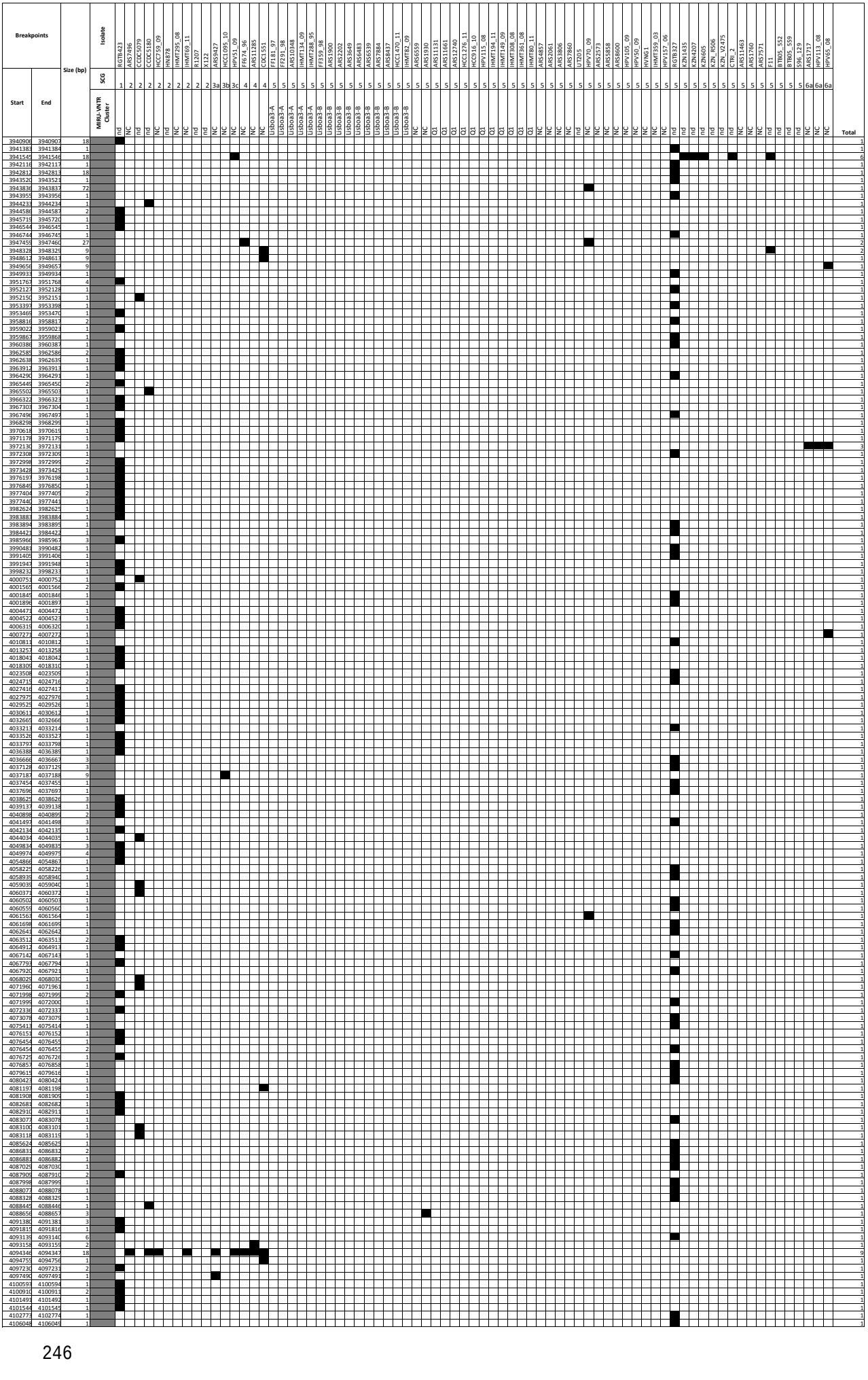


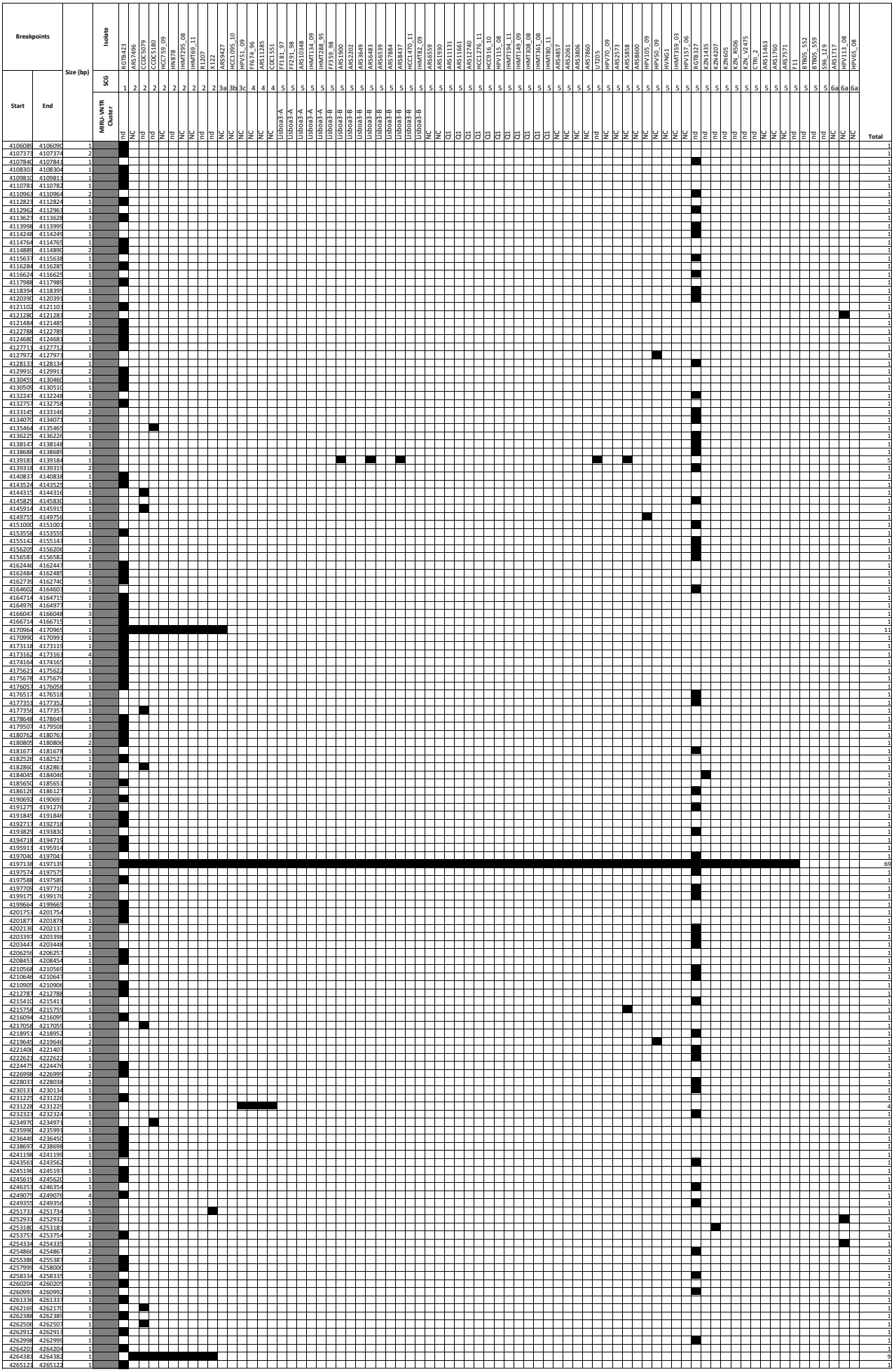
| Breakpoints | | Size (bp) | Isolate | RGTB423 | ARS7406 | CDC5079 | CDC5180 | HCC798_09 | HN878 | IHMT295_08 | IHMT69_11 | R1207 | X122 | ARS9427 | HCC1095_10 | HPV517_09 | FF674_06 | ARS11835 | CDC1551 | FT191_07 | FT233_08 | ARS10348 | IHMT134_09 | IHMT288_05 | FF159_08 | ARS15900 | ARS2202 | ARS3849 | ARS6483 | ARS6539 | ARS7884 | ARS8437 | HCC1470_11 | IHMT82_09 | ARS6559 | ARS1930 | ARS11131 | ARS11661 | ARS12740 | HCC1276_11 | HCC936_10 | HPV115_08 | IHMT194_11 | IHMT149_09 | IHMT308_08 | IHMT361_08 | IHMT80_11 | ARS4857 | ARS2061 | ARS3806 | ARS7860 | UT205 | HPV70_09 | ARS2573 | ARS5858 | ARS8600 | HPV105_09 | HPV50_09 | HVIN61 | IHMT359_03 | HPV157_06 | RGTB327 | KZN1435 | KZN4207 | KZN605 | KZN_R506 | KZN_V2475 | CTRI_2 | ARS11463 | ARS1760 | ARS7571 | F11 | BTB05_552 | BTB05_559 | S96_129 | ARS1717 | HPV113_08 | HPV65_08 | |
|---|---|---


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
FC8157
FC8157_97
FT191_98
ARS10348
HMT134_09
HMT288_95
FT159_98
ARS15900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS9437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVNG1
HMT359_03
HPV157_06
RGTB327
K2N1435
K2N4207
K2N605
K2N_R506
K2N_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
MIRU-VNTR Cluster
nd
NC
nd
nd
NC
nd
NC
NC
nd
nd
NC
NC
NC
NC
NC
NC
U.ubod3-A
U.ubod3-A
U.ubod3-A
U.ubod3-A
U.ubod3-A
U.ubod3-A
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
U.ubod3-B
NC
NC
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
Q1
NC
NC
NC
nd
nd
NC
NC
NC
NC
NC
NC
NC
NC
NC
nd
nd
nd
nd
nd
nd
nd
nd
nd
NC
NC
NC
nd
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT395_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_06
ARS11835
CDC1551
FF191_07
FF191_08
ARS10348
HMT134_09
HMT288_05
FF159_08
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HNB78
HMT295_08
HMT69_11
R1207
X122
ARS9427
HCC1095_10
HPV157_09
FF674_96
ARS11835
CDC1551
FF191_97
FF191_98
ARS10348
HMT194_09
HMT288_95
FF159_98
ARS15900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC758_09
HN878
HMT395_08
HMT69_11
R1207
X122
AR59427
HCC1095_10
FF674_06
AR511265
CDC1551
FF191_07
FF191_08
AR510348
HMT194_09
HMT288_05
FF159_08
AR510900
AR52202
AR53849
AR56483
AR56539
AR57884
AR58437
HCC1470_11
HMT82_09
AR56559
AR51930
AR511131
AR511661
AR512740
HCC1276_11
HCC916_10
HPV115_08
HMT194_11
HMT149_09
HMT308_08
HMT361_08
HMT80_11
AR54857
AR52061
AR53806
AR57860
UT205
HPV70_09
AR52573
AR55858
AR58600
HPV105_09
HPV50_09
HVJNG1
HMT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRL_2
AR511463
AR51760
AR57571
F11
BTB05_552
BTB05_559
S96_129
Total


Breakpoints
Start
End
Size (bp)
Isolate
SCG
MIRU-VNTR Cluster
RGTB423
ARS7496
CDC5079
CDC5180
HCC798_09
HN878
HMIT395_08
HMIT69_11
R1207
X122
ARS9427
HCC1095_10
HPV51_09
FF674_96
ARS11835
CDC1551
FF191_97
FF191_98
ARS10348
HMIT194_09
HMIT288_95
FF159_98
ARS1900
ARS2202
ARS3849
ARS6483
ARS6539
ARS7884
ARS8437
HCC1470_11
HMIT82_09
ARS6559
ARS1930
ARS11131
ARS11661
ARS12740
HCC1276_11
HCC916_10
HPV115_08
HMIT194_11
HMIT149_09
HMIT308_08
HMIT361_08
HMIT80_11
ARS4857
ARS2061
ARS3806
ARS7860
UT205
HPV70_09
ARS2573
ARS5858
ARS8600
HPV105_09
HPV90_09
HVIN61
HMIT359_03
HPV157_06
RGTB327
KZN1435
KZN4207
KZN605
KZN_R506
KZN_V2475
CTRI_2
ARS11463
ARS1760
ARS7571
F11
BTB05_552
BTB05_559
S96_129
ARS1717
HPV113_08
HPV65_08
nd
NC
Total

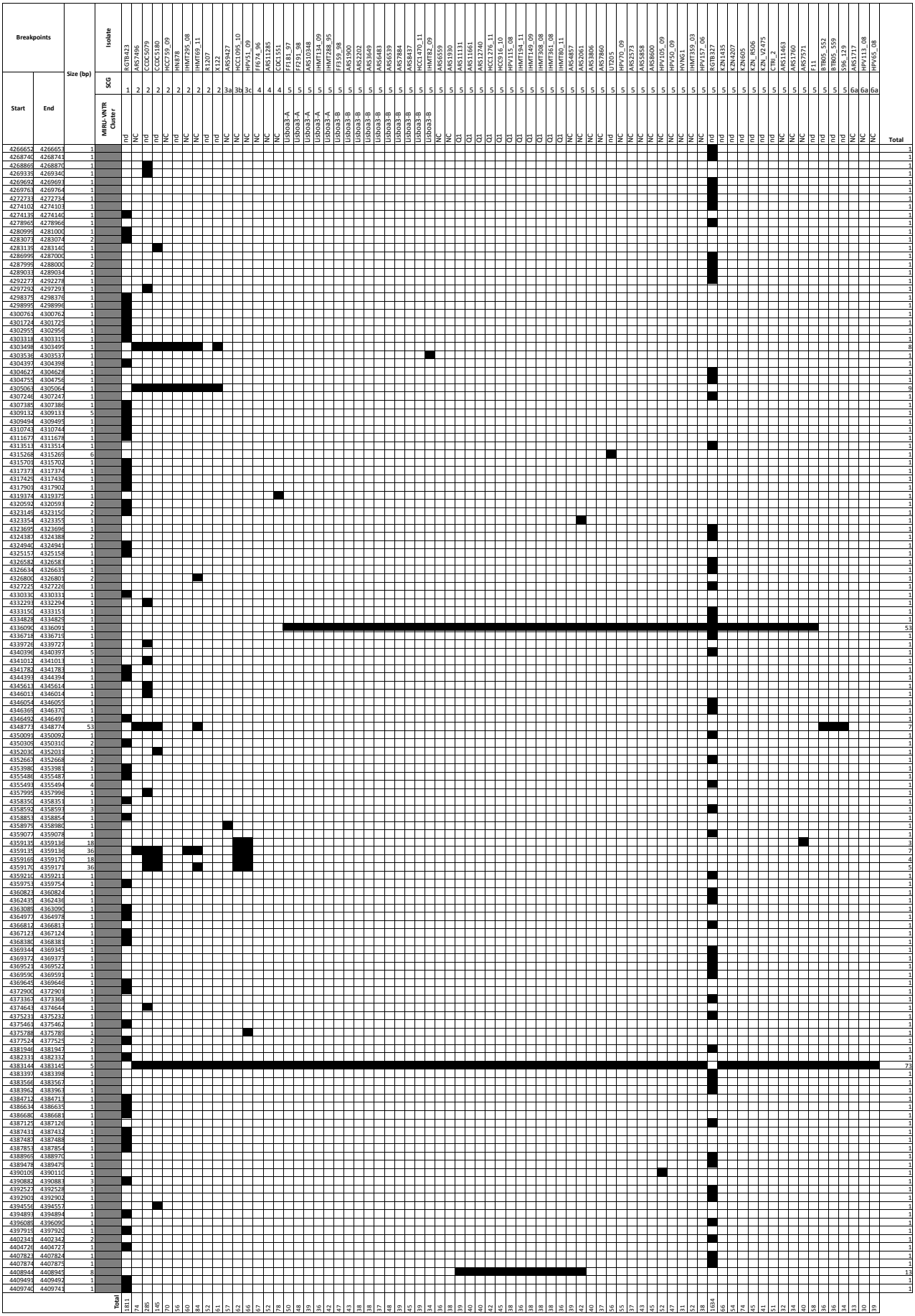
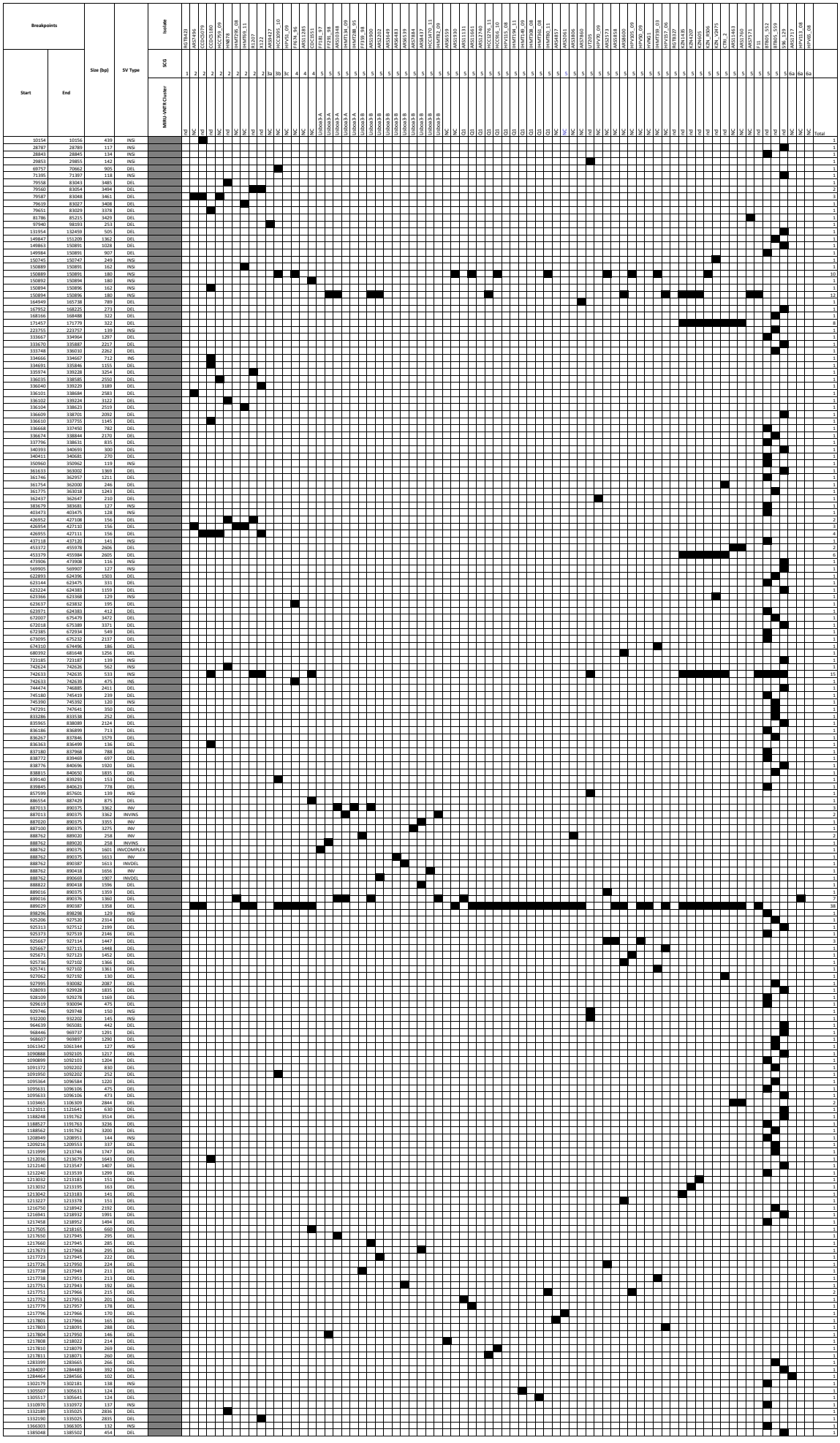
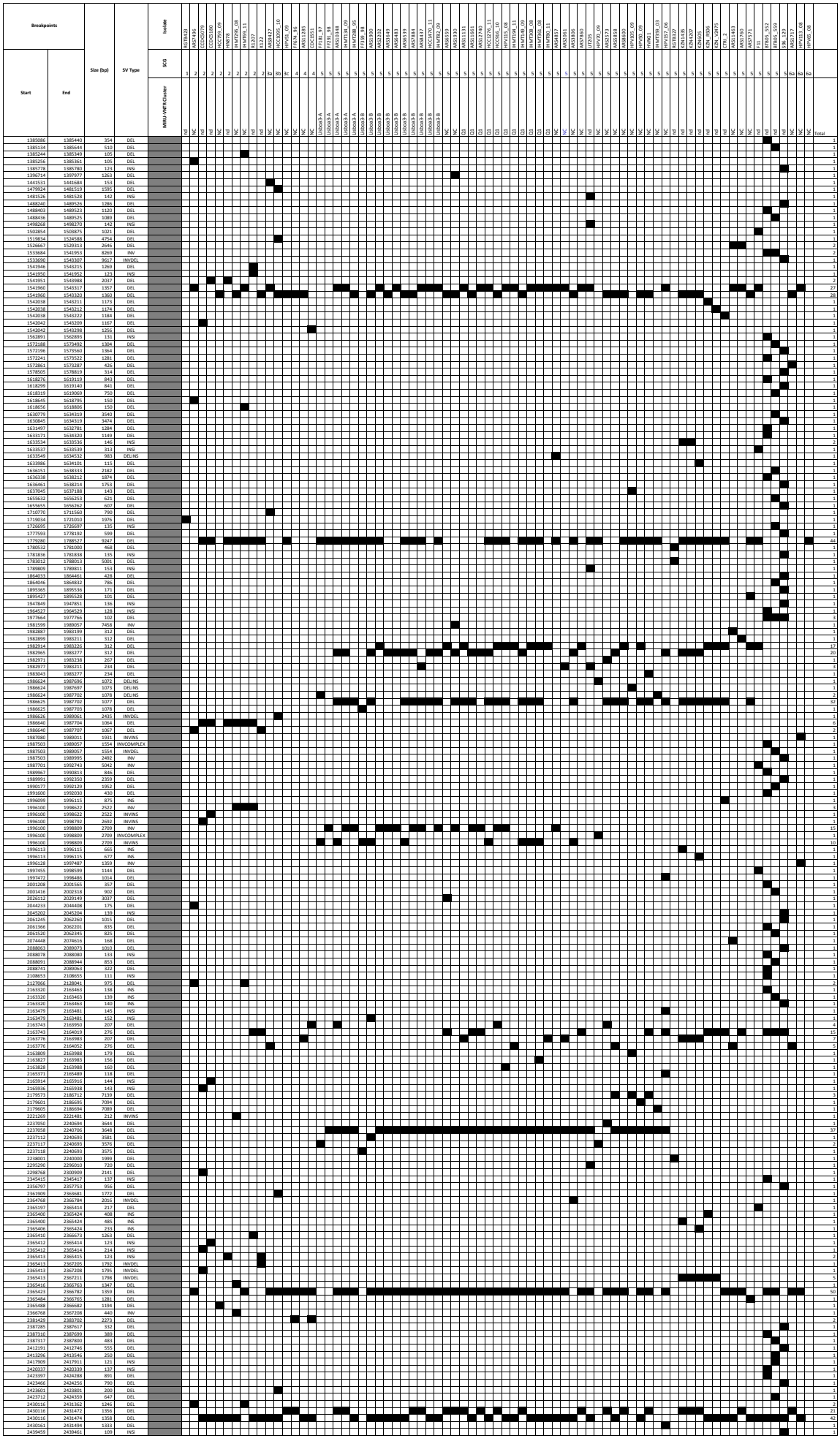
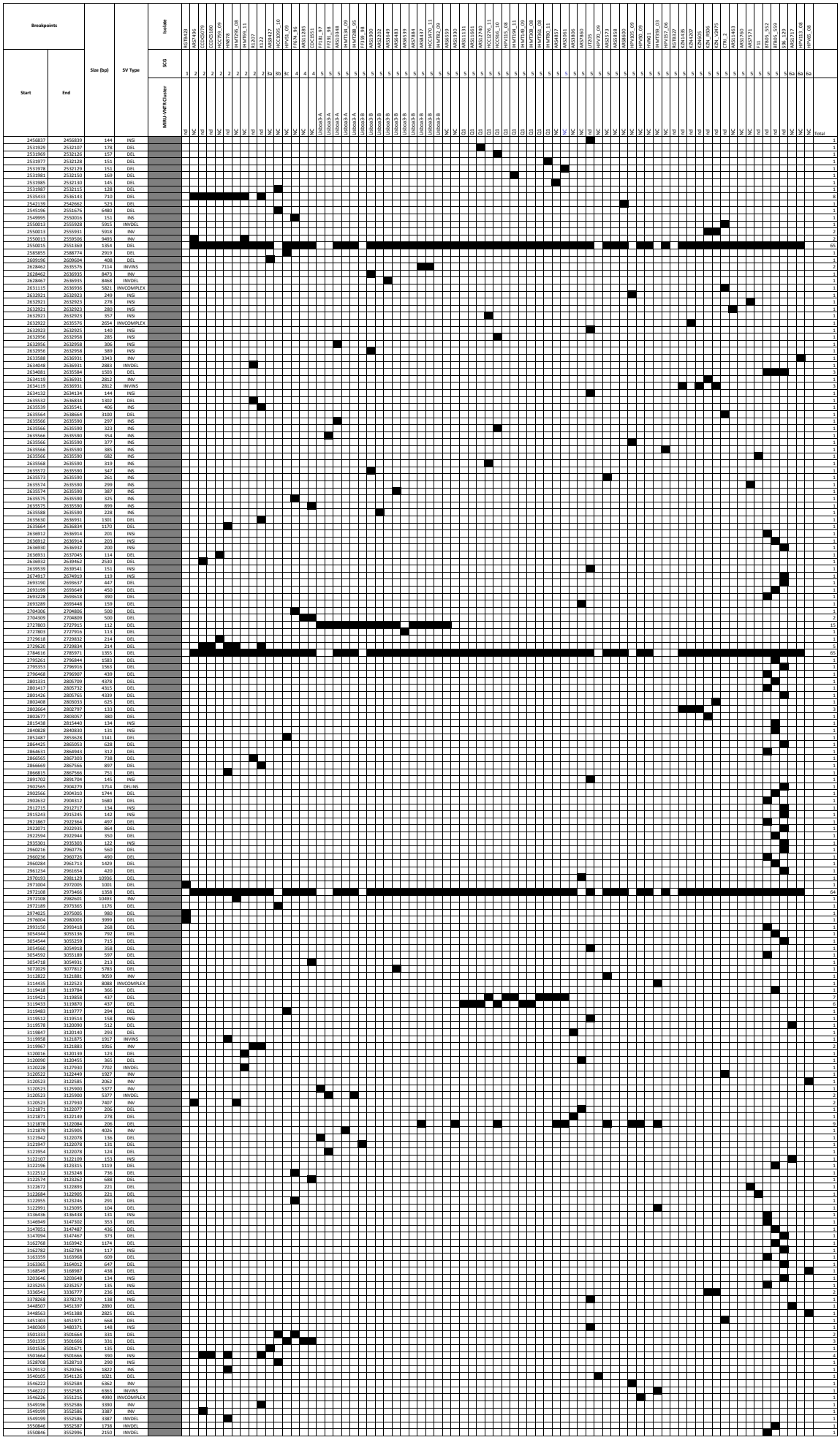


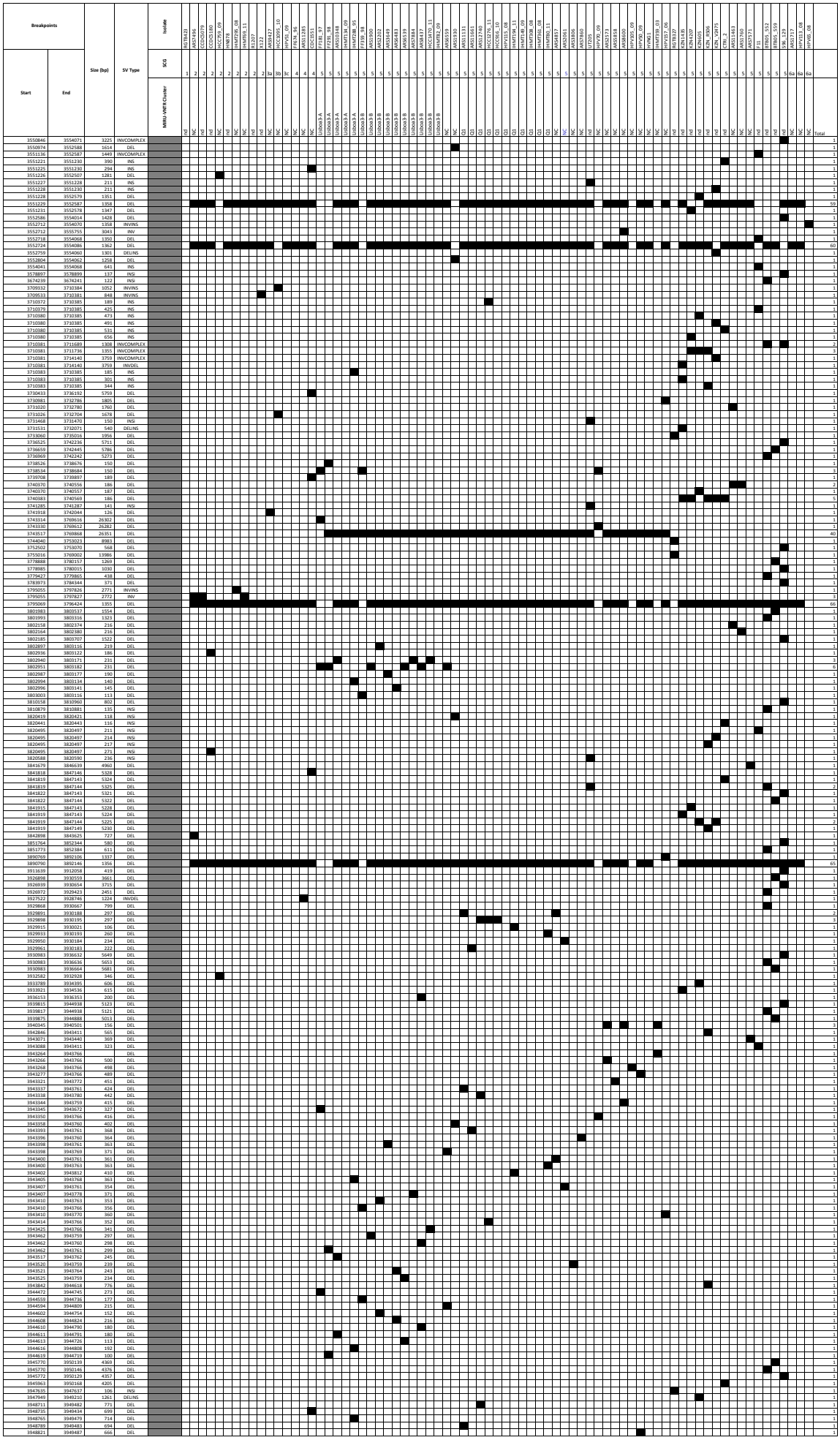
Figure S7.6 - List of short insertions (<100 bp) found among the group of 75 clinical isolates. Black squares indicate insertion detection. MIRU-VNTR cluster indicates the 24-loci MIRU-VNTR cluster of any given isolate, except if non-clustered (NC) or not determined (nd). Line and column totals indicate total column/line count of insertions.

Figure S7.7


Breakpoints
Start
End
Size (bp)
SV Type
Isolate
SCG
MIRU-VNTR Cluster
Total


Breakpoints
Start
End
Size (bp)
SV Type
Isolate
SCG
MIRU-VNTR Cluster
Total


Breakpoints
Start
End
Size (bp)
SV Type
Isolate
SCG
MIRU-VNTR Cluster
Total


Breakpoints
Start
End
Size (bp)
SV Type
Isolate
SCG
MIRU-VNTR Cluster
Total

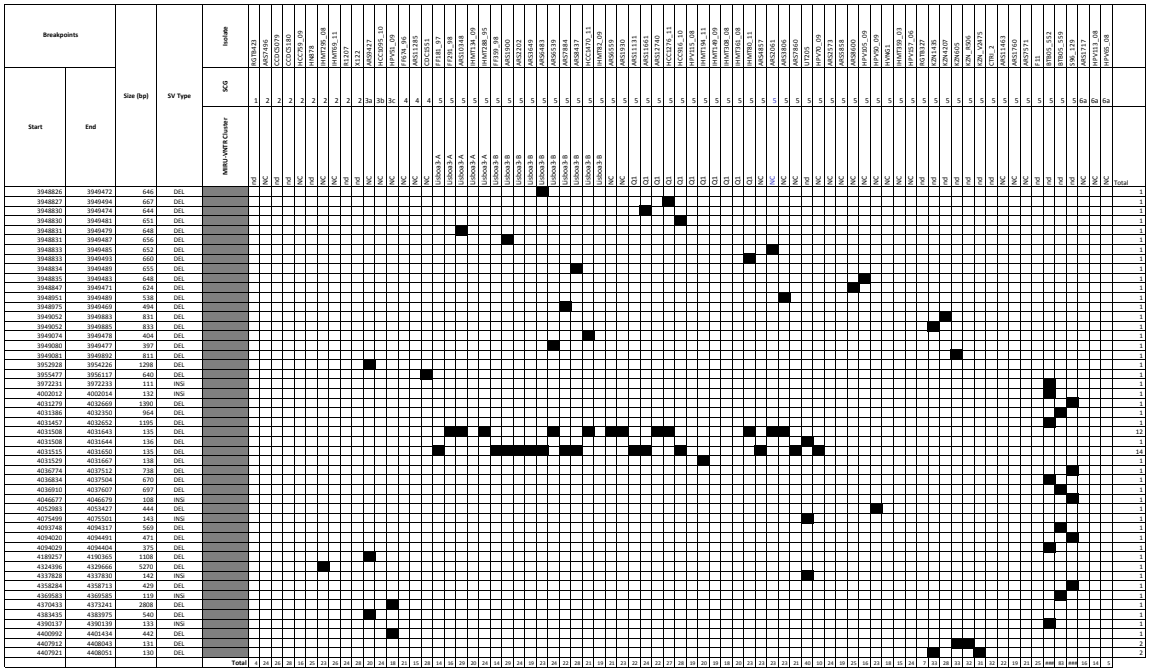
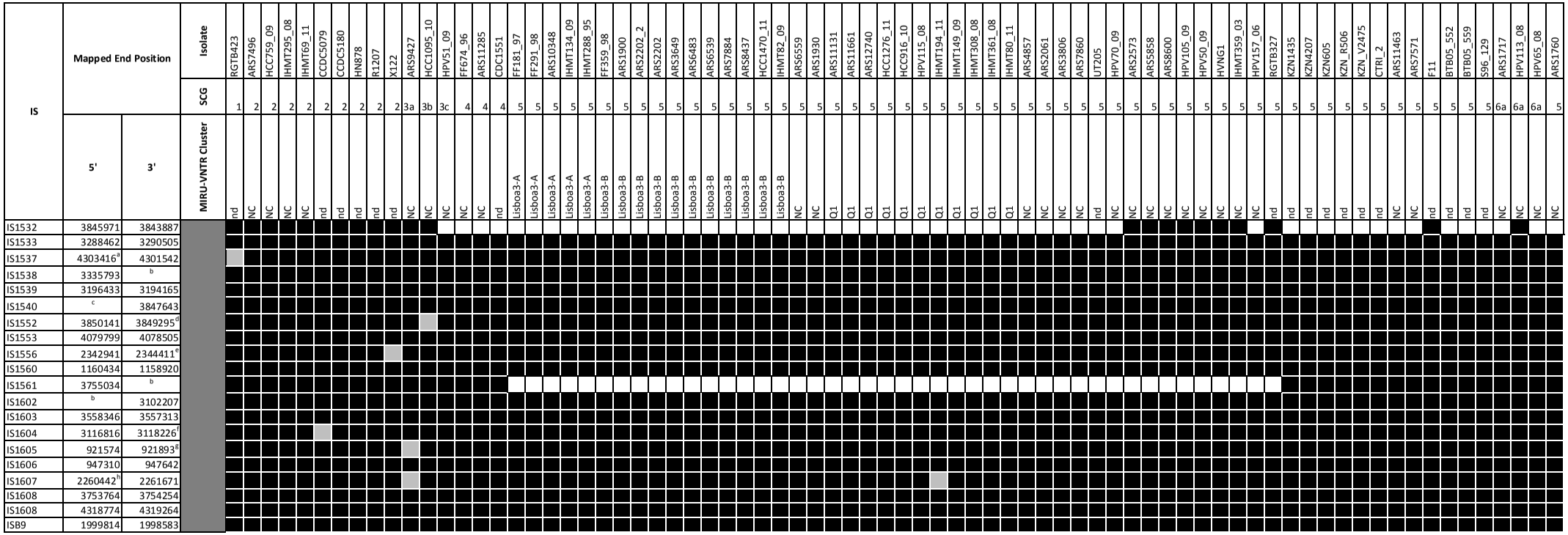


Figure S7.7 - List of SV types found amoing the 75 clinical isolates group using the SVMerge pipeline and excluding copy number gain hits. SV type includes: deletions (DEL); completely (INSi) and incompletely (INS) reconstructed insertions; simple inversions (INV) and complex inversions (INVCOMPLEX); deletions plus insertions (DELINS); and, inversions plus deletions (INVDEL) or plus insertions (INVINS). Black squares are indicative of SV detection. MIRU-VNTR cluster indicates the 24-loci MIRU-VNTR cluster of any given isolate, except if non-clustered (NC) or not determined (nd). Line and column totals indicate total column/line count of SVs.



| IS | Mapped End Position | | Isolate | RGTB423 | ARS7496 | HCC759_09 | IHMT295_08 | IHMT69_11 | CCDC5079 | CCDC5180 | HN878 | R1207 | X122 | ARS9427 | HCC1095_10 | HPV51_09 | FF674_96 | ARS11285 | CDC1551 | FF181_97 | FF291_98 | ARS10348 | IHMT134_09 | IHMT288_95 | FF359_98 | ARS1900 | ARS2202_2 | ARS2202 | ARS3649 | ARS6483 | ARS6539 | ARS7884 | ARS8437 | HCC1470_11 | IHMT82_09 | ARS6559 | ARS1930 | ARS11131 | ARS11661 | ARS12740 | HCC1276_11 | HCC916_10 | HPV115_08 | IHMT194_11 | IHMT149_09 | IHMT308_08 | IHMT361_08 | IHMT80_11 | ARS4857 | ARS2061 | ARS3806 | ARS7860 | UT205 | HPV70_09 | ARS2573 | ARS5858 | ARS8600 | HPV105_09 | HPV50_09 | HVNG1 | IHMT359_03 | HPV157_06 | RGTB327 | KZN1435 | KZN4207 | KZN605 | KZN_R506 | KZN_V2475 | CTRL_2 | ARS11463 | ARS7571 | F11 | BTB05_552 | BTB05_559 | S96_129 | ARS1717 | HPV113_08 | HPV65_08 | ARS1760 |
|---|---|---


Position
Mapped End
Confidence
Chain
ORF
Isolate
SGS
MIRU-VNTR Cluster
Total

Figure S7.9 - Mapped positions of IS6110 found across the genomes of the 75 analyzed *M. tuberculosis* clinical isolates in relation to *M. tuberculosis* H37Rv. Each mapped position shown refers to a IS6110 end from which the genomic position of insertion was deduced, referred on the Mapped End column. Chain column shows the chain coding for IS6110 copy in question and consequently, its orientation. Confidence column corresponds to the quality/confidence level classification explained in the Materials and Methods section. ORF column shows: the affected ORF in case of an intragenic insertion site; intergenic if the site is intergenic and mapped IS is not on the proper orientation to exert a putative upregulatory effect on an ORF located downstream of the IS 3' end; or, the preffix up indicating that the mapped IS is upstream and in the same orientation of a downstream ORF, followed by a number indicating the distance to the downstream ORF and, followed by the ORF designation, gene or feature designation. Black squares indicate IS6110 copies mapped at both 5' and 3' end; grey squares indicate IS6110 copies that only the mapped end indicated in the Mapped End column was mapped; and, yellow squares indicate IS6110 copies on which the only mapped end is the other end than the one indicated in the Mapped End column. MIRU-VNTR cluster indicates the 24-loci MIRU-VNTR cluster of any given isolate, except if non-clustered (NC) or not determined (nd). Column and line totals account for the the number of IS6110 copies mapped on each line and column, respectively. Isolates highlighted in red and blue belong to Lisboa3 and Q1 clades, respectively.

Table S7.1- Number of mutations categorized by structural and functional effect type found along specified branches of the Lisboa3 subtree.

Lisboa 3	**Tree Branch**						
	A-B	**B-B1**	**B-C**	**C-D**	**D-D1**	**D-E**	**E-E1**
Resistance acquisition[a]	INH STP	RIF EMB FQ KAN AMK CAP			RIF	RIF	EMB PZA FQ KAN
Structural type							
SNPs	23	237	2	2	80	48	187
INDELS		78			3		5
Mutational Effect							
Synonymous	10	95	2	2	31	20	73
Missense	12	113			44	23	94
Nonsense		3				1	2
Frameshift		38			3		
Start Lost		1			1		1
Stop Lost		1					
Codon Deletion		6					2
Codon Insertion		3					
Codon Change + Codon Deletion		3					1
Codon Change + Codon Insertion		3					
rRNA mutation							
Intergenic	1	49			4	4	19

[a]INH, isoniazid; HL, High-level resistance; STP, streptomycin; RIF, rifampicin; EMB, ethambutol; PZA, pyrazinamide; KAN, kanamycin; AMK; amikacin; CAP, capreomycin; FQ, fluoroquinolone.

Table S7.2 – Number of mutations categorized by structural and functional effect type found along specified branches of the Q1 subtree.

Q1	**Tree Branch**					
	A-B	**B-C**	**C-D**	**D-E**	**E-F**	**F-G**
Resistance acquisition[a]	INH	INH (HL) STP	RIF EMB PZA		KAN AMK CAP	FQ
Structural type						
SNPs	36	38	5	1	1	1
INDELS		4	1			2
Mutational Effect						
Synonymous	11	17				
Missense	19	21	5	1		2
Nonsense						
Frameshift		3				
Start Lost						
Stop Lost						
Codon Deletion						
Codon Insertion						
Codon Change + Codon Deletion						
Codon Change + Codon Insertion						
rRNA mutation					1	
Intergenic	6	3	1			1

[a]INH, isoniazid; HL, High-level resistance; STP, streptomycin; RIF, rifampicin; EMB, ethambutol; PZA, pyrazinamide; KAN, kanamycin; AMK; amikacin; CAP, capreomycin; FQ, fluoroquinolone.

Table S7.3 - Mutations found to be acquired along node-delimited branches in the Lisboa3 subtree. Position, Reference Sequence and Mutated Sequence are derived from the VCF format.

Branch	Position	Reference Sequence	Mutated Sequence	Type	Codon Change[a]	Affected ORF	Mutation
A-B							
	405812	T	C	Missense	atA/atG	Rv0338c	I10M
	740561	C	T	Synonymous	ccG/ccA	lipG	P193
	781687	A	G	Missense	aAg/aGg	rpsL	K43R
	1057412	A	G	Missense	cTa/cCa	Rv0947c	L40P
	1153137	G	A	Missense	Gcg/Acg	kdpA	A376T
	1673425	C	T	Intergenic			
	1674481	T	G	Missense	Tcg/Gcg	inhA	S94A
	1706439	G	A	Synonymous	Ctg/Ttg	Rv1514c	L53
	1885422	C	T	Missense	tCg/tTg	pks8	S1240L
	2123086	C	A	Synonymous	ccG/ccT	lldD2	P22
	2255294	T	C	Missense	aTc/aCc	otsB1	I1098T
	2264189	C	T	Synonymous	caC/caT	Rv2017	H64
	2817383	C	T	Synonymous	ctG/ctA	accD1	L364
	3256167	C	T	Synonymous	taC/taT	ppsC	Y161
	3385982	C	T	Synonymous	cgG/cgA	Rv3026c	R32
	3431653	G	A	Missense	Gat/Aat	Rv3067	D76N
	3434043	G	A	Missense	Gtg/Atg	ccrB	V118M
	3540581	A	G	Synonymous	caT/caC	hpx	H55
	3544710	T	C	Synonymous	ccA/ccG	mesT	P197
	3730361	C	T	Missense	gGc/gAc	PPE54	G2192D
	3883690	G	A	Missense	Gcg/Acg	Rv3466	A56T
	4034894	G	A	Synonymous	ccG/ccA	lpqF	P181
	4104091	C	T	Missense	Gcc/Acc	dppC	A129T
B-B1							
	6817	G	A	Synonymous	aaG/aaA	gyrB	K565
	7572	T	C	Missense	Tcg/Ccg	gyrA	S91P
	21795	G	A	Missense	Ccg/Tcg	ppp	P463S
	42967	G	C	Synonymous	ccC/ccG	mtc28	P133
	49478	CCGTTGC	CC	Frameshift		Rv0045c	
	63771	C	T	Missense	cCg/cTg	Rv0059	P191L
	71336	G	C	Missense	cGc/cCc	Rv0064	R906P
	74059	C	T	Synonymous	aaG/aaA	icd2	K151
	79504	TCGGTGGACCCGGTGGACCCGGTGGACCCGGTGG ACCCGGTGGACCCGGTGGAC	TCGGTGGACCCGGTGGACCCGGTGGACCCGGTGG ACCCGGTGGAC	Codon deletion	ccggtggac/-	Rv0071	PVD22-
	101727	G	A	Missense	gGa/gAa	ctpA	G382E
	103756	G	T	Missense	cCc/cAc	Rv0094c	P303H
	103896	C	G	Missense	gaG/gaC	Rv0094c	E256D
	104944	G	A	Missense	gCc/gTc	Rv0095c	A91V
	125830	GAAAA	GAAAAA	Frameshift		ctpI	
	131174	TGG	TGGG	Intergenic			
	154283	T	C	Missense	Tcg/Ccg	Rv0127	S18P
	194681	G	C	Synonymous	ctC/ctG	Rv0165c	L86
	196642	C	T	Synonymous	aaC/aaT	fadD5	N550
	208287	A	C	Missense	gAa/gCa	Rv0176	E279A
	223942	T	C	Missense	Tcg/Ccg	Rv0192	S127P
	234496	C	CGT	Frameshift		Rv0197	
	293704	CTTTTT	CTTTT	Intergenic			
	316129	C	G	Missense	gGc/gCc	Rv0264c	G96A
	333892	G	C	Missense	Cgc/Ggc	PE_PGRS3	R807G
	337820	G	A	Synonymous	ggC/ggT	PE_PGRS4	G418
	338876	G	A	Synonymous	agC/agT	PE_PGRS4	S66
	338960	T	C	Synonymous	gcA/gcG	PE_PGRS4	A38
	338963	T	C	Synonymous	acA/acG	PE_PGRS4	T37
	338984	C	T	Synonymous	gcG/gcA	PE_PGRS4	A30
	340132	G	A	Missense	Gaa/Aaa	PPE3	E257K
	373282	TAA	TA	Frameshift		PPE6	
	411118	CCGCGGCGCG	CCGCG	Frameshift		iniA	
	424320	TCC	TCCC	Frameshift		PPE7	
	427310	TTGCCGAGGTTTGCACTGCCGAGGTTTGCACTGCC GAGGTTTGCACTGCCGAGGTTTGCACTGCC	TTGCCGAGGTTTGCACTGCCGAGGTTTGCACTGCC GAGGTTTGCACTGCC	Codon deletion	ggcagtgcaaacctc/-	PPE8	GSANL2436-
	467497	CGGG	CGGGG	Frameshift		PPE9	
	467526	C	G	Missense	gGg/gCg	PPE9	G159A
	467546	G	C	Missense	gaC/gaG	PPE9	D152E
	467557	A	C	Missense	Ttg/Gtg	PPE9	L149V
	467564	A	C	Missense	caT/caG	PPE9	H146Q
	467590	T	C	Missense	Acg/Gcg	PPE9	T138A
	467621	T	G	Synonymous	ggA/ggC	PPE9	G127
	467638	G	T	Missense	Cag/Aag	PPE9	Q122K
	475977	C	T	Synonymous	gcC/gcT	Rv0397	A54
	489935	G	C	Missense	cGg/cCg	pks6	R1402P
	510831	TCCGGGGGGCGCACCGGGGG	TCCGGGGG	Codon deletion	tcccccggtgcgccc/tcc	ctpH	SPGAP1491S
	580773	GGGGGCACCACCCGCTTGCGGGGGAG	GG	Intergenic			
	583956	G	C	Synonymous	ggC/ggG	Rv0493c	G245
	589536	G	A	Synonymous	ctG/ctA	Rv0499	L118
	608037	A	C	Missense	cAc/cCc	Rv0515	H496P
	623472	A	G	Missense	gAc/gGc	PE_PGRS6	D227G
	632330	G	T	Synonymous	cgG/cgT	Rv0539	R196
	637319	G	A	Missense	Cct/Tct	pitA	P49S
	637922	C	T	Synonymous	caG/caA	Rv0546c	Q16
	664249	C	T	Synonymous	caG/caA	Rv0571c	Q190
	669398	T	C	Synonymous	caA/caG	Rv0575c	Q116
	672491	C	G	Synonymous	ggG/ggC	PE_PGRS7	G1142
	673238	A	G	Synonymous	caT/caC	PE_PGRS7	H893
	685461	C	G	Synonymous	gcC/gcG	yrbE2A	A111
	685608	T	C	Synonymous	ctT/ctC	yrbE2A	L160
	686972	T	C	Missense	tTc/tCc	mce2A	F51S
	709226	G	A	Missense	aCc/aTc	Rv0613c	T97I
	761155	C	T	Missense	tCg/tTg	rpoB	S450L
	766823	A	C	Missense	Aag/Cag	rpoC	K1152Q
	769962	GCCC	GCC	Frameshift		end	
	773809	G	A	Synonymous	gaG/gaA	echA4	E229
	781922	A	G	Synonymous	aaA/aaG	rpsL	K121
	836272	A	G	Missense	gAg/gGg	PE_PGRS9	E191G
	836291	A	G	Synonymous	ggA/ggG	PE_PGRS9	G197
	836658	A	G	Missense	Acc/Gcc	PE_PGRS9	T320A
	839334	A	G	Missense	aAa/aGa	PE_PGRS10	K295R
	839348	A	G	Missense	Agc/Ggc	PE_PGRS10	S300G
	839949	C	G	Missense	gCg/gGg	PE_PGRS10	A500G
	853579	C	G	Missense	gCc/gGc	phoR	A395G
	854252	GCCCCCCCCC	GCCCCCCC	Intergenic			
	859131	C	CA	Frameshift		cyp123	
	859383	C	T	Missense	atG/atA	cyp123	M230I
	874835	CC	CCGC	Frameshift		ptrBa	
	889014	AAGGAGG	A	Intergenic			
	921813	C	G	Missense	gCa/gGa	Rv0829	A80G
	927385	A	G	Synonymous	ggA/ggG	PE_PGRS13	G675
	947429	T	A	Missense	Tcg/Acg	Rv0850	S40T
	947430	C	A	Nonsense	tCg/tAg	Rv0850	S40*
	955524	A	G	Missense	Agc/Ggc	fadA	S150G
	968426	AGCCGGGTTGGCCGG	AGCCGGGTTGGCCGGGTTGGCCGG	Codon insertion	-/CCGGCCAAC	PE_PGRS15	-602PAN
	976896	TTG	T	Frameshift		PPE13	
	986463	G	C	Intergenic			
	990001	G	C	Missense	Ccc/Gcc	Rv0890c	P866A
	993346	A	C	Missense	gTt/gGt	Rv0891c	V37G

1010204 CGG	CGGG	Frameshift		Rv0907	
1041129 C	T	Intergenic			
1061676 GTGCTG	GTG	Intergenic			
1074558 G	A	Missense	cCc/cTc	lprP	P186L
1079927 C	A	Synonymous	acC/acA	ctpV	T417
1093406 A	G	Synonymous	gtT/gtC	PE_PGRS17	V317
1093886 C	A	Synonymous	ggG/ggT	PE_PGRS17	G157
1093928 G	A	Synonymous	aaC/aaT	PE_PGRS17	N143
1096470 G	C	Intergenic			
1103249 C	T	Synonymous	gcC/gcT	Rv0987	A236
1107917 G	T	Missense	caC/caA	Rv0990c	H61Q
1150321 TGCCCGACAGCAAGGCC	TGCC	Frameshift		kdpD	
1159161 GCCCACC	GCC	Frameshift		Rv1034c	
1163134 T	C	Synonymous	ggA/ggG	PE8	G81
1165521 TA	TAA	Intergenic			
1168715 CTT	CTTT	Frameshift		Rv1046c	
1199547 G	A	Missense	cCa/cTa	Rv1075c	P275L
1212076 G	A	Missense	Ggc/Agc	PE_PGRS21	G173S
1218737 G	T	Missense	Ggc/Tgc	PE_PGRS22	G757C
1224367 T	C	Intergenic			
1244700 T	C	Synonymous	Tta/Cta	zwf1	L332
1276321 T	G	Missense	caA/caC	Rv1148c	Q476H
1276322 T	G	Missense	cAa/cCa	Rv1148c	Q476P
1277869 GT	GTT	Intergenic			
1307598 C	G	Missense	tGc/tCc	fadH	C210S
1313337 A	AG	Intergenic			
1328687 G	C	Missense	Ccc/Gcc	Rv1186c	P207A
1365837 CGGGG	CGGGGG	Intergenic			
1373170 G	C	Missense	cCg/cGg	Rv1230c	P343R
1374936 CAA	CA	Frameshift		Rv1232c	
1404169 T	G	Missense	aaA/aaC	cyp130	K145N
1423805 G	C	Synonymous	acC/acG	Rv1273c	T82
1440469 C	G	Synonymous	ccC/ccG	cysN	P521
1447774 C	T	Synonymous	Ctg/Ttg	argS	L466
1452071 C	A	Synonymous	ggC/ggA	thrB	G25
1457144 C	T	Missense	Cgc/Tgc	hemK	R194C
1484708 A	C	Missense	Agc/Cgc	Rv1321	S144R
1533241 G	C	Synonymous	acC/acG	PPE19	T131
1533583 G	A	Synonymous	taC/taT	PPE19	Y17
1547125 T	C	Missense	Aca/Gca	Rv1374c	T136A
1563717 C	T	Synonymous	gtC/gtT	mihF	V8
1586249 C	T	Synonymous	caG/caA	Rv1410c	Q506
1588456 A	G	Synonymous	cgT/cgC	lprG	R9
1625347 CGGG	CGGGG	Frameshift		opcA	
1631066 GGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCC GGTGCC	GGCCGCCGGTGCCGCCGGTGCCGCCGGTGCC	Codon deletion	ggcaccggc/-	PE_PGRS27	GTG1175-
1636826 C	A	Synonymous	ggG/ggT	PE_PGRS28	G468
1638211 T	C	Missense	Acc/Gcc	PE_PGRS28	T7A
1638212 T	C	Synonymous	gtA/gtG	PE_PGRS28	V6
1639594 C	A	Missense	cCa/cAa	Rv1453	P405Q
1645802 T	C	Missense	Aaa/Gaa	Rv1459c	K113E
1656646 TAA	TAAA	Frameshift		PE_PGRS29	
1692141 A	C	Synonymous	atA/atC	Rv1501	I84
1719322 G	A	Synonymous	ctG/ctA	Rv1524	L199
1728837 A	G	Intergenic			
1753519 GCCCC	GCCCCC	Frameshift		fadD11.1	
1759252 G	T	Synonymous	tcG/tcT	frdA	S524
1789650 C	T	Missense	Gcc/Acc	Rv1588c	A63T
1789654 A	G	Synonymous	ctT/ctC	Rv1588c	L61
1789671 C	T	Missense	Gcg/Acg	Rv1588c	A56T
1798355 G	A	Missense	gGt/gAt	Rv1597	G21D
1804409 C	A	Missense	cCa/cAa	impA	P124Q
1863700 G	C	Synonymous	ggC/ggG	PE_PGRS30	G561
1894300 GG	GGTCTTGCCGCG	Frameshift		Rv1668c	
1918690 TTGT	TTGTGT	Frameshift		tlyA	
1933988 G	A	Intergenic			
1944402 T	C	Missense	gTc/gCc	Rv1716	V276A
1944642 CT	C	Frameshift		Rv1717	
1967237 C	A	Missense	cGa/cTa	Rv1739c	R134L
1983313 T	G	Missense	aAc/aCc	PPE24	N488T
1986623 GTT	GTTT	Frameshift		Rv1754c	
2045310 A	G	Synonymous	atT/atC	PE_PGRS32	I511
2052035 G	T	Missense	Gtg/Ttg	PPE33	V252L
2055271 A	G	Missense	cTa/cCa	Rv1812c	L30P
2058257 T	G	Start lost	tTg/tGg	Rv1816	L1W
2074565 C	G	Intergenic			
2075094 C	T	Missense	gCg/gTg	Rv1830	A85V
2094911 ACAGCGTCAGCGTCAG	ACAGCGTCAG	Codon deletion	actgacgct/act	gnd1	TDA88T
2096186 A	G	Synonymous	acT/acC	Rv1846c	T138
2109523 CGG	CGGG	Intergenic			
2137521 AC	ACTCCGATCACC	Frameshift		Rv1888c	
2143328 G	C	Missense	Gtg/Ctg	Rv1895	V270L
2197148 C	A	Missense	aCc/aAc	Rv1945	T387N
2207525 C	T	Intergenic			
2207591 T	TC	Intergenic			
2211826 A	G	Synonymous	aaA/aaG	mce3C	K67
2231879 T	C	Missense	gTc/gCc	Rv1988	V67A
2268725 A	AG	Frameshift		Rv2024c	
2269780 T	C	Missense	gAc/gGc	Rv2024c	D154G
2278442 C	G	Synonymous	cgG/cgC	Rv2030c	R15
2300205 A	AC	Frameshift		pks12	
2300555 A	G	Synonymous	gaT/gaC	pks12	D2144
2335075 A	G	Missense	gAa/gGa	Rv2078	E6G
2340621 C	G	Missense	cCc/cGc	Rv2082	P638R
2345037 C	A	Synonymous	ctC/ctA	pknJ	L209
2348446 C	G	Missense	ttC/ttG	Rv2090	F358L
2357268 TGCCGCCGCCGCCGCCGCCG	TGCCGCCGCCGCCGCCG	Intergenic			
2365413 AT	ATT	Intergenic			
2368564 TAA	TA	Intergenic			
2382085 AGTG	AG	Frameshift		PPE37	
2415656 G	C	Missense	Cgg/Ggg	murD	R250G
2424925 A	G	Intergenic			
2434782 A	G	Intergenic			
2439204 A	G	Intergenic			
2439401 A	G	Missense	Tac/Cac	Rv2177c	Y183H
2468058 C	T	Intergenic			
2470886 AAACTTAGCCGCGA	AA	Codon change and Codon deletion	aaacttagccgcgag/aag	Rv2206	KLSRE89K
2499726 G	A	Missense	Gat/Aat	Rv2226	D299N
2503625 C	T	Synonymous	ctG/ctA	Rv2230c	L328
2509140 G	C	Missense	tCc/tGc	cobD	S79C
2523205 GCG	GCGCCG	Intergenic			
2534562 GGAG	GG	Frameshift		Rv2262c	
2573756 C	A	Intergenic			
2617632 A	AT	Intergenic			
2626018 T	C	Missense	gAg/gGg	esxO	E52G
2626026 A	G	Synonymous	gcT/gcC	esxO	A49
2626095 C	G	Synonymous	gcG/gcC	esxO	A26
2626155 C	T	Synonymous	caG/caA	esxO	Q6
2654371 G	A	Synonymous	aaC/aaT	hrcA	N241
2695378 C	G	Missense	gGc/gCc	cysW	G141A
2704884 AC	ACAGCGACCATATCGCCGAGCTC	Codon insertion	-/AGCGACCATATCGCCGAGCTC	Rv2407	-64SDHIAEL
2715342 C	T	Intergenic			

	2734074 T	C	Missense	gTc/gCc	rbsK	V282A
	2830525 C	A	Missense	aCg/aAg	Rv2513	T122K
	2865760 A	G	Missense	Acc/Gcc	Rv2542	T211A
	2868659 C	G	Synonymous	gcC/gcG	Rv2547	A18
	2881597 AGGGGGG	AGGGGG	Frameshift		Rv2561	
	2945167 G	T	Intergenic			
	2955571 A	G	Intergenic			
	3011401 A	G	Stop lost	Tga/Cga	Rv2694c	*123R
	3028101 ACC	ACCC	Frameshift		Rv2715	
	3054081 A	G	Synonymous	ggA/ggG	PE_PGRS47	G56
	3073868 T	C	Missense	Acc/Gcc	thyA	T202A
	3077039 C	A	Missense	gGc/gTc	PPE43	G347V
	3080795 A	G	Missense	cTc/cCc	Rv2771c	L80P
	3131469 TTGT	TTGTCGGCGATGT	Codon insertion	-/ACATCGCCG	Rv2823c	-101TSP
	3133536 T	C	Missense	Aaa/Gaa	Rv2825c	K2E
	3158935 G	C	Missense	Cgc/Ggc	Rv2850c	R374G
	3178445 C	G	Intergenic			
	3179242 ACGGCGGCGGCG	ACGGCGGCG	Codon deletion	gacgcc/gac	Rv2867c	DA20D
	3190145 TCCCCCC	TCCCCC	Frameshift		Rv2880c	
	3191358 G	A	Synonymous	atC/atT	cdsA	I88
	3226181 A	C	Synonymous	cgT/cgG	ffh	R35
	3232815 A	G	Intergenic			
	3247853 C	T	Synonymous	gcC/gcT	ppsA	A803
	3247856 G	C	Synonymous	cgG/cgC	ppsA	R804
	3247874 G	A	Synonymous	cgG/cgA	ppsA	R810
	3247883 T	C	Synonymous	agT/agC	ppsA	S813
	3302324 AGCGAG	A	Frameshift		fadD29	
	3312632 C	T	Nonsense	tgG/tgA	Rv2959c	W69*
	3354896 C	T	Synonymous	gaG/gaA	serA1	E58
	3363338 A	G	Intergenic			
	3377271 G	C	Missense	Ctg/Gtg	PPE46	L325V
	3377275 G	C	Synonymous	gcC/gcG	PPE46	A323
	3377281 T	C	Synonymous	ggA/ggG	PPE46	G321
	3377305 T	C	Synonymous	gtA/gtG	PPE46	V313
	3377347 C	A	Synonymous	ggG/ggT	PPE46	G299
	3379718 T	C	Synonymous	ggA/ggG	PPE47	G245
	3379742 T	C	Synonymous	gtA/gtG	PPE47	V237
	3379751 A	C	Synonymous	gtT/gtG	PPE47	V234
	3379784 C	A	Synonymous	ggG/ggT	PPE47	G223
	3402816 C	T	Missense	gGa/gAa	serB2	G116E
	3415180 ACACCTAGGGGGTGGCACCTAGGGGGTGGCAC	ACACCTAGGGGGTGGCAC	Intergenic			
	3418328 T	G	Missense	gAt/gCt	Rv3057c	D112A
	3418330 G	A	Synonymous	caC/caT	Rv3057c	H111
	3473996 GA	GAA	Intergenic			
	3480375 GAAA	GAA	Frameshift		Rv3113	
	3484012 T	G	Synonymous	gcT/gcG	cysA3	A13
	3503895 C	T	Missense	cCg/cTg	Rv3137	P168L
	3504930 C	T	Synonymous	Ctg/Ttg	pflA	L246
	3510642 T	C	Missense	Agc/Ggc	PPE52	S226G
	3519294 GGCC	GGCCGCC	Codon change and Codon insertion	cac/cGCCac	nuoH	H6RH
	3554787 G	A	Missense	gGa/gAa	Rv3189	G49E
	3561155 G	A	Missense	gCc/gTc	Rv3193c	A673V
	3580636 CTTT	CTT	Frameshift		lipV	
	3590686 GT	GCT	Intergenic			
	3604821 G	C	Synonymous	gcG/gcC	Rv3228	A32
	3661530 C	T	Synonymous	caG/caA	birA	Q161
	3689523 G	T	Nonsense	tgC/tgA	lpdA	C472*
	3730741 G	A	Synonymous	ggC/ggT	PPE54	G2065
	3736261 A	G	Synonymous	ggT/ggC	PPE54	G225
	3739913 G	A	Synonymous	acC/acT	PE_PGRS50	T954
	3769550 G	A	Synonymous	taC/taT	Rv3355c	Y86
	3779671 CGGCAACGGTGGCAACGGTGGCAACGGTG	CGGCAACGGTGGCAACGGTGGCAACGGTGGCAAC GGTG	Codon change and Codon insertion	gcc/gGCAACGGTGcc	PE_PGRS51	A378GNGA
	3794884 G	A	Intergenic			
	3840393 A	G	Missense	aTg/aCg	alr	M343T
	3847378 G	C	Missense	Gac/Cac	PPE59	D72H
	3847380 C	A	Missense	gaC/gaA	PPE59	D72E
	3859376 C	T	Missense	gGg/gAg	Rv3439c	G96E
	3862472 GAA	GA	Intergenic			
	3883707 T	C	Synonymous	ctT/ctC	Rv3466	L61
	3883711 G	A	Missense	Gcc/Acc	Rv3466	A63T
	3884906 A	G	Missense	Aaa/Gaa	Rv3467	K315E
	3890778 CTGA	C	Intergenic			
	3892671 A	G	Synonymous	gtT/gtC	kgtP	V350
	3930476 ACGGCG	ACGGCGGCACCGGCG	Codon change and Codon insertion	gtt/gGCACCGGCGtt	PE_PGRS53	V1305GTGV
	3934699 G	A	Missense	aGc/aAc	PE_PGRS54	S1232N
	3940802 A	G	Missense	Aac/Gac	PE_PGRS55	N396D
	3942481 C	G	Missense	gCg/gGg	PE_PGRS56	A253G
	3942640 T	C	Missense	aTt/aCt	PE_PGRS56	I306T
	3998059 AGGCGGCGGCGGC	AGGCGGCGGC	Codon change and Codon deletion	gcggcc/gcc	PPE64	AA30A
	4037283 T	G	Synonymous	ggA/ggC	PE_PGRS59	G256
	4038287 G	A	Synonymous	aaC/aaT	clpC1	N806
	4093879 TGGGGG	TGGGG	Frameshift		PE_PGRS60	
	4095001 CG	C	Frameshift		Rv3655c	
	4120983 A	G	Intergenic			
	4135112 G	A	Missense	atG/atA	Rv3693	M129I
	4187817 A	G	Missense	gAt/gGt	Rv3737	D40G
	4197138 CTT	CTTT	Intergenic			
	4198611 CGGGGG	CGGGG	Intergenic			
	4218350 T	C	Missense	aAg/aGg	Rv3773c	K159R
	4247702 C	A	Missense	Ccg/Acg	embB	P397T
	4264219 T	G	Synonymous	ggA/ggC	Rv3802c	G49
	4287370 A	G	Missense	cAg/cGg	Rv3822	Q217R
	4293072 G	A	Missense	Ctt/Ttt	papA1	L35F
	4320050 A	G	Intergenic			
	4336090 A	AT	Intergenic			
	4338595 GC	G	Intergenic			
	4338732 G	A	Intergenic			
	4353201 CCCGCCGCCGCC	CCCGCCGCC	Codon change and Codon deletion	ccgcca/cca	Rv3876	PP67P
	4373475 C	G	Synonymous	gtG/gtC	Rv3889c	V52
	4373496 C	G	Synonymous	gtG/gtC	Rv3889c	V45
	4382275 G	T	Missense	Caa/Aaa	Rv3896c	Q193K
	4383144 CC	CCGGGGC	Frameshift		Rv3897c	
	4385177 G	C	Missense	cCa/cGa	Rv3899c	P68R
	4395964 C	A	Synonymous	acC/acA	Rv3909	T591
	4400660 ACCC	ACC	Frameshift		sigM	
	4410200 G	A	Synonymous	ctC/ctT	Rv3922c	L72
B-C						
	1567562 C	T	Synonymous	aaC/aaT	metK	N246
	3379757 A	C	Synonymous	gcT/gcG	PPE47	A232
D-C						
	2998574 G	T	Synonymous	ggC/ggA	dxs1	G465
	3261702 G	A	Synonymous	gcG/gcA	ppsC	A2006
D-D1						
	42967 G	C	Synonymous	ccC/ccG	mtc28	P133
	63771 C	T	Missense	cCg/cTg	Rv0059	P191L
	71336 G	C	Missense	cGc/cCc	Rv0064	R906P
	71585 CG	CGAG	Intergenic			
	103756 G	T	Missense	cCc/cAc	Rv0094c	P303H

	194681	G	C	Synonymous	ctC/ctG	Rv0165c	L86
	208287	A	C	Missense	gAa/gCa	Rv0176	E279A
	467557	A	C	Missense	Ttg/Gtg	PPE9	L149V
	467564	A	C	Missense	caT/caG	PPE9	H146Q
	467621	T	G	Synonymous	ggA/ggC	PPE9	G127
	467638	G	T	Missense	Cag/Aag	PPE9	Q122K
	475977	C	T	Synonymous	gcC/gcT	Rv0397	A54
	583956	G	C	Synonymous	ggC/ggG	Rv0493c	G245
	632330	G	T	Synonymous	cgG/cgT	Rv0539	R196
	637922	C	T	Synonymous	caG/caA	Rv0546c	Q16
	669398	T	C	Synonymous	caA/caG	Rv0575c	Q116
	673238	A	G	Synonymous	caT/caC	PE_PGRS7	H893
	709226	G	A	Missense	aCc/aTc	Rv0613c	T97I
	761110	A	T	Missense	gAc/gTc	rpoB	D435V
	773809	G	A	Synonymous	gaG/gaA	echA4	E229
	836538	A	G	Missense	Aac/Gac	PE_PGRS9	N280D
	836658	A	G	Missense	Acc/Gcc	PE_PGRS9	T320A
	853579	C	G	Missense	gCc/gGc	phoR	A395G
	859383	C	T	Missense	atG/atA	cyp123	M230I
	921813	C	G	Missense	gCa/gGa	Rv0829	A80G
	947430	C	A	Nonsense	tCg/tAg	Rv0850	S40*
	1041129	C	T	Intergenic			
	1074558	G	A	Missense	cCc/cTc	lprP	P186L
	1103249	C	T	Synonymous	gcC/gcT	Rv0987	A236
	1218737	G	T	Missense	Ggc/Tgc	PE_PGRS22	G757C
	1307598	C	G	Missense	tGc/tCc	fadH	C210S
	1447774	C	T	Synonymous	Ctg/Ttg	argS	L466
	1457144	C	T	Missense	Cgc/Tgc	hemK	R194C
	1484708	A	C	Missense	Agc/Cgc	Rv1321	S144R
	1588456	A	G	Synonymous	cgT/cgC	lprG	R9
	1692141	A	C	Synonymous	atA/atC	Rv1501	I84
	1697055	A	G	Missense	cTa/cCa	Rv1507c	L123P
	1719322	G	A	Synonymous	ctG/ctA	Rv1524	L199
	1789671	C	T	Missense	Gcg/Acg	Rv1588c	A56T
	1798355	G	A	Missense	gGt/gAt	Rv1597	G21D
	1944402	T	C	Missense	gTc/gCc	Rv1716	V276A
	1986625	TCATC	TTGCATC	Frameshift		Rv1754c	
	2045310	A	G	Synonymous	atT/atC	PE_PGRS32	I511
	2052035	G	T	Missense	Gtg/Ttg	PPE33	V252L
	2055271	A	G	Missense	cTa/cCa	Rv1812c	L30P
	2058257	T	G	Start lost	tTg/tGg	Rv1816	L1W
	2143328	G	C	Missense	Gtg/Ctg	Rv1895	V270L
	2269780	T	C	Missense	gAc/gGc	Rv2024c	D154G
	2278442	C	G	Synonymous	cgG/cgC	Rv2030c	R15
	2300205	ATGT	ACTGT	Frameshift		pks12	
	2348446	C	G	Missense	ttC/ttG	Rv2090	F358L
	2401883	C	T	Intergenic			
	2509140	G	C	Missense	tCc/tGc	cobD	S79C
	2626018	T	C	Missense	gAg/gGg	esxO	E52G
	2748308	A	G	Missense	Tcc/Ccc	valS	S640P
	2811885	G	A	Missense	gCg/gTg	pdhA	A71V
	2868659	C	G	Synonymous	gcC/gcG	Rv2547	A18
	2955571	A	G	Intergenic			
	3073868	T	C	Missense	Acc/Gcc	thyA	T202A
	3077039	C	A	Missense	gGc/gTc	PPE43	G347V
	3080795	A	G	Missense	cTc/cCc	Rv2771c	L80P
	3247856	G	C	Synonymous	cgG/cgC	ppsA	R804
	3247883	T	C	Synonymous	agT/agC	ppsA	S813
	3354896	C	T	Synonymous	gaG/gaA	serA1	E58
	3377305	T	C	Synonymous	gtA/gtG	PPE46	V313
	3379718	T	C	Synonymous	ggA/ggG	PPE47	G245
	3379742	T	C	Synonymous	gtA/gtG	PPE47	V237
	3379751	A	C	Synonymous	gtT/gtG	PPE47	V234
	3379763	G	A	Synonymous	gcC/gcT	PPE47	A230
	3484012	T	G	Synonymous	gcT/gcG	cysA3	A13
	3503895	C	T	Missense	cGg/cTg	Rv3137	P168L
	3510642	T	C	Missense	Agc/Ggc	PPE52	S226G
	3561155	G	A	Missense	gCc/gTc	Rv3193c	A673V
	3604821	G	C	Synonymous	gcG/gcC	Rv3228	A32
	3892671	A	G	Synonymous	gtT/gtC	kgtP	V350
	3934542	T	G	Missense	Tct/Gct	PE_PGRS54	S1180A
	3942640	T	C	Missense	aTt/aCt	PE_PGRS56	I306T
	4187817	A	G	Missense	gAt/gGt	Rv3737	D40G
	4218350	T	C	Missense	aAg/aGg	Rv3773c	K159R
	4287370	A	G	Missense	cAg/cGg	Rv3822	Q217R
	4373496	C	G	Synonymous	gtG/gtC	Rv3889c	V45
	4385177	G	C	Missense	cCa/cGa	Rv3899c	P68R
	4395964	C	A	Synonymous	acC/acA	Rv3909	T591
D-E							
	2378	G	A	Synonymous	ttG/ttA	dnaN	L109
	6817	G	A	Synonymous	aaG/aaA	gyrB	K565
	196642	C	T	Synonymous	aaC/aaT	fadD5	N550
	223942	T	C	Missense	Tcg/Ccg	Rv0192	S127P
	589536	G	A	Synonymous	ctG/ctA	Rv0499	L118
	608037	A	C	Missense	cAc/cCc	Rv0515	H496P
	685608	T	C	Synonymous	ctT/ctC	yrbE2A	L160
	761155	C	T	Missense	tCg/tTg	rpoB	S450L
	947429	T	A	Missense	Tcg/Acg	Rv0850	S40T
	955524	A	G	Missense	Agc/Ggc	fadA	S150G
	1079927	C	A	Synonymous	acC/acA	ctpV	T417
	1093406	A	G	Synonymous	gtT/gtC	PE_PGRS17	V317
	1107917	G	T	Missense	caC/caA	Rv0990c	H61Q
	1199547	G	A	Missense	cCa/cTa	Rv1075c	P275L
	1533583	G	A	Synonymous	taC/taT	PPE19	Y17
	1547125	T	C	Missense	Aca/Gca	Rv1374c	T136A
	1645802	T	C	Missense	Aaa/Gaa	Rv1459c	K113E
	1689562	A	G	Missense	tTc/tCc	Rv1498c	F120S
	1759252	G	T	Synonymous	tcG/tcT	frdA	S524
	1789650	C	T	Missense	Gcc/Acc	Rv1588c	A63T
	1789654	A	G	Synonymous	ctT/ctC	Rv1588c	L61
	1967237	C	A	Missense	cGa/cTa	Rv1739c	R134L
	1983313	T	G	Missense	aAc/aCc	PPE24	N488T
	2096186	A	G	Synonymous	acT/acC	Rv1846c	T138
	2207525	C	T	Intergenic			
	2231879	T	C	Missense	gTc/gCc	Rv1988	V67A
	2335075	A	G	Missense	gAa/gGa	Rv2078	E6G
	2345037	C	A	Synonymous	ctC/ctA	pknJ	L209
	2468058	C	T	Intergenic			
	2499726	G	A	Missense	Gat/Aat	Rv2226	D299N
	2503625	C	T	Synonymous	ctG/ctA	Rv2230c	L328
	2654371	G	A	Synonymous	aaC/aaT	hrcA	N241
	2695378	C	G	Missense	gGc/gCc	cysW	G141A
	2830525	C	A	Missense	aCg/aAg	Rv2513	T122K
	2962129	C	A	Missense	Gta/Tta	PE_PGRS46	V105L
	3377271	G	C	Missense	Ctg/Gtg	PPE46	L325V
	3377275	G	C	Synonymous	gcC/gcG	PPE46	A323
	3379712	G	C	Synonymous	gcC/gcG	PPE47	A247
	3504930	C	T	Synonymous	Ctg/Ttg	pflA	L246
	3654567	G	A	Missense	Gcg/Acg	Rv3272	A374T
	3661530	C	T	Synonymous	caG/caA	birA	Q161
	3689523	G	T	Nonsense	tgC/tgA	lpdA	C472*

	3794884	G	A	Intergenic			
	3859376	C	T	Missense	gGg/gAg	Rv3439c	G96E
	4038287	G	A	Synonymous	aaC/aaT	clpC1	N806
	4135112	G	A	Missense	atG/atA	Rv3693	M129I
	4338732	G	A	Intergenic			
	4410200	G	A	Synonymous	ctC/ctT	Rv3922c	L72
E-E1							
	7582	A	G	Missense	gAc/gGc	gyrA	D94G
	21795	G	A	Missense	Ccg/Tcg	ppp	P463S
	42967	G	C	Synonymous	ccC/ccG	mtc28	P133
	63771	C	T	Missense	cCg/cTg	Rv0059	P191L
	71336	G	C	Missense	cGc/cCc	Rv0064	R906P
	74059	C	T	Synonymous	aaG/aaA	icd2	K151
	101727	G	A	Missense	gGa/gAa	ctpA	G382E
	103756	G	T	Missense	cCc/cAc	Rv0094c	P303H
	104944	G	A	Missense	gCc/gTc	Rv0095c	A91V
	154283	T	C	Missense	Tcg/Ccg	Rv0127	S18P
	179886	G	A	Synonymous	gaC/gaT	PE2	D337
	194681	G	C	Synonymous	ctC/ctG	Rv0165c	L86
	208287	A	C	Missense	gAa/gCa	Rv0176	E279A
	316129	C	G	Missense	gGc/gCc	Rv0264c	G96A
	333892	G	C	Missense	Cgc/Ggc	PE_PGRS3	R807G
	337820	G	A	Synonymous	ggC/ggT	PE_PGRS4	G418
	338020	A	C	Missense	Tgc/Ggc	PE_PGRS4	C352G
	338100	T	C	Missense	aAc/aGc	PE_PGRS4	N325S
	338876	G	A	Synonymous	agC/agT	PE_PGRS4	S66
	338960	T	C	Synonymous	gcA/gcG	PE_PGRS4	A38
	340132	G	A	Missense	Gaa/Aaa	PPE3	E257K
	467526	C	G	Missense	gGg/gCg	PPE9	G159A
	467546	G	C	Missense	gaC/gaG	PPE9	D152E
	467557	A	C	Missense	Ttg/Gtg	PPE9	L149V
	467564	A	C	Missense	caT/caG	PPE9	H146Q
	467590	T	C	Missense	Acg/Gcg	PPE9	T138A
	467621	T	G	Synonymous	ggA/ggC	PPE9	G127
	467638	G	T	Missense	Cag/Aag	PPE9	Q122K
	475977	C	T	Synonymous	gcC/gcT	Rv0397	A54
	489935	G	C	Missense	cGg/cCg	pks6	R1402P
	490949	C	G	Missense	tCc/tGc	fgd1	S56C
	583956	G	C	Synonymous	ggC/ggG	Rv0493c	G245
	623472	A	G	Missense	gAc/gGc	PE_PGRS6	D227G
	623508	C	G	Missense	gCc/gGc	PE_PGRS6	A239G
	632330	G	T	Synonymous	cgG/cgT	Rv0539	R196
	637319	G	A	Missense	Cct/Tct	pitA	P49S
	637922	C	T	Synonymous	caG/caA	Rv0546c	Q16
	664249	C	T	Synonymous	caG/caA	Rv0571c	Q190
	669398	T	C	Synonymous	caA/caG	Rv0575c	Q116
	672491	C	G	Synonymous	ggG/ggC	PE_PGRS7	G1142
	673238	A	G	Synonymous	caT/caC	PE_PGRS7	H893
	685461	C	G,T	Synonymous	gcC/gcG	yrbE2A	A111
	686972	T	C	Missense	tTc/tCc	mce2A	F51S
	709226	G	A	Missense	aCc/aTc	Rv0613c	T97I
	773809	G	A	Synonymous	gaG/gaA	echA4	E229
	781922	A	G,T	Missense	aaA/aaT	rpsL	K121N
	836272	A	G,T	Missense	gAg/gGg	PE_PGRS9	E191G
	836291	A	G	Synonymous	ggA/ggG	PE_PGRS9	G197
	836426	A	C	Synonymous	ctA/ctC	PE_PGRS9	L242
	836454	A	G	Missense	Act/Gct	PE_PGRS9	T252A
	836538	A	G	Missense	Aac/Gac	PE_PGRS9	N280D
	836658	A	G	Missense	Acc/Gcc	PE_PGRS9	T320A
	839334	A	G	Missense	aAa/aGa	PE_PGRS10	K295R
	839348	A	G	Missense	Agc/Ggc	PE_PGRS10	S300G
	839949	C	G,T	Missense	gCg/gGg	PE_PGRS10	A500G
	853579	C	G	Missense	gCc/gGc	phoR	A395G
	859383	C	T	Missense	atG/atA	cyp123	M230I
	921813	C	G	Missense	gCa/gGa	Rv0829	A80G
	927110	A	G	Missense	Agt/Ggt	PE_PGRS13	S584G
	927385	A	G	Synonymous	ggA/ggG	PE_PGRS13	G675
	947430	C	A	Nonsense	tCg/tAg	Rv0850	S40*
	986463	G	C	Intergenic			
	990001	G	C	Missense	Ccc/Gcc	Rv0890c	P866A
	993346	A	C	Missense	gTt/gGt	Rv0891c	V37G
	1041129	C	T	Intergenic			
	1074558	G	A	Missense	cCc/cTc	lprP	P186L
	1093886	C	A	Synonymous	ggG/ggT	PE_PGRS17	G157
	1103249	C	T	Synonymous	gcC/gcT	Rv0987	A236
	1163134	T	C,A	Synonymous	ggA/ggG	PE8	G81
	1189431	TGCCGCCGCCGCCGCCG	TGCCGCCGCCGCCG	Codon deletion	aacggc/aac	PE_PGRS19	NG326N
	1212076	G	A	Missense	Ggc/Agc	PE_PGRS21	G173S
	1218737	G	T	Missense	Ggc/Tgc	PE_PGRS22	G757C
	1224367	T	C	Intergenic			
	1244700	T	C	Synonymous	Tta/Cta	zwf1	L332
	1276321	T	G	Missense	caA/caC	Rv1148c	Q476H
	1276322	T	G	Missense	cAa/cCa	Rv1148c	Q476P
	1307598	C	G	Missense	tGc/tCc	fadH	C210S
	1328687	G	C	Missense	Ccc/Gcc	Rv1186c	P207A
	1373170	G	C,A	Missense	cCg/cGg	Rv1230c	P343R
	1404169	T	G	Missense	aaA/aaC	cyp130	K145N
	1423805	G	C	Synonymous	acC/acG	Rv1273c	T82
	1440469	C	G	Synonymous	ccC/ccG	cysN	P521
	1447774	C	T	Synonymous	Ctg/Ttg	argS	L466
	1452071	C	A	Synonymous	ggC/ggA	thrB	G25
	1457144	C	T	Missense	Cgc/Tgc	hemK	R194C
	1484708	A	C	Missense	Agc/Cgc	Rv1321	S144R
	1563717	C	T	Synonymous	gtC/gtT	mihF	V8
	1586249	C	T	Synonymous	caG/caA	Rv1410c	Q506
	1588456	A	G	Synonymous	cgT/cgC	lprG	R9
	1631066	GGCCGCCGGTGCCGCCGGTGCCGCCGGTGCCGCC GGTGCC	GGCCGCCGGTGCCGCCGGTGCCGCCGGTGCC	Codon deletion	ggcaccggc/-	PE_PGRS27	GTG1175-
	1634580	T	C	Synonymous	ttA/ttG	PE_PGRS27	L16
	1634581	A	G	Missense	tTa/tCa	PE_PGRS27	L16S
	1636826	C	A	Synonymous	ggG/ggT	PE_PGRS28	G468
	1638211	T	C	Missense	Acc/Gcc	PE_PGRS28	T7A
	1638212	T	C	Synonymous	gtA/gtG	PE_PGRS28	V6
	1639594	C	A	Missense	cCa/cAa	Rv1453	P405Q
	1692141	A	C	Synonymous	atA/atC	Rv1501	I84
	1719322	G	A	Synonymous	ctG/ctA	Rv1524	L199
	1728837	A	G	Intergenic			
	1789671	C	T	Missense	Gcg/Acg	Rv1588c	A56T
	1798355	G	A	Missense	gGt/gAt	Rv1597	G21D
	1804409	C	A	Missense	cCa/cAa	impA	P124Q
	1863700	G	C,A	Synonymous	ggC/ggG	PE_PGRS30	G561
	1933988	G	A	Intergenic			
	1944402	T	C	Missense	gTc/gCc	Rv1716	V276A
	2045310	A	G	Synonymous	atT/atC	PE_PGRS32	I511
	2052035	G	T	Missense	Gtg/Ttg	PPE33	V252L
	2055271	A	G	Missense	cTa/cCa	Rv1812c	L30P
	2058257	T	G	Start lost	tTg/tGg	Rv1816	L1W
	2074565	C	G	Intergenic			
	2143328	G	C	Missense	Gtg/Ctg	Rv1895	V270L
	2211826	A	G	Synonymous	aaA/aaG	mce3C	K67
	2269780	T	C	Missense	gAc/gGc	Rv2024c	D154G

	2278442	C	G	Synonymous	cgG/cgC	Rv2030c	R15
	2288883	A	G	Missense	cTg/cCg	pncA	L120P
	2300237	A	G	Synonymous	gcT/gcC	pks12	A2250
	2300552	T	G	Synonymous	ccA/ccC	pks12	P2145
	2300555	A	G	Synonymous	gaT/gaC	pks12	D2144
	2340621	C	G	Missense	cCc/cGc	Rv2082	P638R
	2348446	C	G	Missense	ttC/ttG	Rv2090	F358L
	2415656	G	C,A	Missense	Cgg/Ggg	murD	R250G
	2424925	A	G	Intergenic			
	2439204	A	G,T	Intergenic			
	2509140	G	C	Missense	tCc/tGc	cobD	S79C
	2573756	C	A	Intergenic			
	2626018	T	C	Missense	gAg/gGg	esxO	E52G
	2626026	A	G	Synonymous	gcT/gcC	esxO	A49
	2626155	C	T	Synonymous	caG/caA	esxO	Q6
	2715342	C	T	Intergenic			
	2734074	T	C	Missense	gTc/gCc	rbsK	V282A
	2784613	A	G	Intergenic			
	2865760	A	G	Missense	Acc/Gcc	Rv2542	T211A
	2868659	C	G	Synonymous	gcC/gcG	Rv2547	A18
	2955571	A	G	Intergenic			
	3054081	A	G	Synonymous	ggA/ggG	PE_PGRS47	G56
	3080795	A	G	Missense	cTc/cCc	Rv2771c	L80P
	3133536	T	C	Missense	Aaa/Gaa	Rv2825c	K2E
	3158935	G	C	Missense	Cgc/Ggc	Rv2850c	R374G
	3178445	C	G	Intergenic			
	3191358	G	A	Synonymous	atC/atT	cdsA	I88
	3226181	A	C	Synonymous	cgT/cgG	ffh	R35
	3232815	A	G	Intergenic			
	3247853	C	T	Synonymous	gcC/gcT	ppsA	A803
	3247874	G	A	Synonymous	cgG/cgA	ppsA	R810
	3247877	T	C	Synonymous	ttT/ttC	ppsA	F811
	3247883	T	C	Synonymous	agT/agC	ppsA	S813
	3312632	C	T	Nonsense	tgG/tgA	Rv2959c	W69*
	3354896	C	T	Synonymous	gaG/gaA	serA1	E58
	3363338	A	G	Intergenic			
	3377305	T	C	Synonymous	gtA/gtG	PPE46	V313
	3377347	C	A	Synonymous	ggG/ggT	PPE46	G299
	3379718	T	C	Synonymous	ggA/ggG	PPE47	G245
	3379751	A	C	Synonymous	gtT/gtG	PPE47	V234
	3379763	G	A	Synonymous	gcC/gcT	PPE47	A230
	3379784	C	A	Synonymous	ggG/ggT	PPE47	G223
	3402816	C	T	Missense	gGa/gAa	serB2	G116E
	3418328	T	G	Missense	gAt/gCt	Rv3057c	D112A
	3418330	G	A	Synonymous	caC/caT	Rv3057c	H111
	3484012	T	G	Synonymous	gcT/gcG	cysA3	A13
	3503895	C	T	Missense	cCg/cTg	Rv3137	P168L
	3510642	T	C	Missense	Agc/Ggc	PPE52	S226G
	3554787	G	A	Missense	gGa/gAa	Rv3189	G49E
	3560178	C	T	Intergenic			
	3561155	G	A	Missense	gCc/gTc	Rv3193c	A673V
	3604821	G	C	Synonymous	gcG/gcC	Rv3228	A32
	3739913	G	A	Synonymous	acC/acT	PE_PGRS50	T954
	3769550	G	A	Synonymous	taC/taT	Rv3355c	Y86
	3796409	TTCA	T	Intergenic			
	3841083	A	C	Missense	cTg/cGg	alr	L113R
	3883707	T	C	Synonymous	ctT/ctC	Rv3466	L61
	3883711	G	A	Missense	Gcc/Acc	Rv3466	A63T
	3890778	CTGA	C	Intergenic			
	3892671	A	G	Synonymous	gtT/gtC	kgtP	V350
	3928784	CCGGCGGGGCCGGCGG	CCGGCGGGGCCGGCGGTAACGGCGGGGCCGGCG G	Codon change and Codon insertion	ggc/ggTAACGGCGGGGCCGGCGGc	PE_PGRS53	G744GNGGAGG
	3934542	T	G	Missense	Tct/Gct	PE_PGRS54	S1180A
	3934699	G	A	Missense	aGc/aAc	PE_PGRS54	S1232N
	3940802	A	G	Missense	Aac/Gac	PE_PGRS55	N396D
	3942481	C	G	Missense	gCg/gGg	PE_PGRS56	A253G
	3942640	T	C	Missense	aTt/aCt	PE_PGRS56	I306T
	3943124	C	T	Synonymous	ggC/ggT	PE_PGRS56	G467
	4037283	T	G	Synonymous	ggA/ggC	PE_PGRS59	G256
	4187817	A	G	Missense	gAt/gGt	Rv3737	D40G
	4218350	T	C	Missense	aAg/aGg	Rv3773c	K159R
	4247429	A	G	Missense	Atg/Gtg	embB	M306V
	4287370	A	G	Missense	cAg/cGg	Rv3822	Q217R
	4293072	G	A	Missense	Ctt/Ttt	papA1	L35F
	4320050	A	G	Intergenic			
	4373475	C	G	Synonymous	gtG/gtC	Rv3889c	V52
	4373496	C	G	Synonymous	gtG/gtC	Rv3889c	V45
	4382275	G	T	Missense	Caa/Aaa	Rv3896c	Q193K
	4385177	G	C	Missense	cCa/cGa	Rv3899c	P68R
	4395964	C	A	Synonymous	acC/acA	Rv3909	T591
E-F							
	49478	CCGTTGC	CC	Frameshift		Rv0045c	
	79504	TCGGTGGACCCGGTGGACCCGGTGGACCCGGTGG ACCCGGTGGACCCGGTGGAC	TCGGTGGACCCGGTGGACCCGGTGGACCCGGTGG ACCCGGTGGAC	Codon deletion	ccggtggac/-	Rv0071	PVD22-
	125830	GAAAA	GAAAAA	Frameshift		ctpI	
	131174	TGG	TGGG	Intergenic			
	234496	C	CGT	Frameshift		Rv0197	
	293704	CTTTTT	CTTTT	Intergenic			
	373282	TAA	TA	Frameshift		PPE6	
	411118	CCGCGGCGCG	CCGCG	Frameshift		iniA	
	424320	TCC	TCCC	Frameshift		PPE7	
	427310	TTGCCGAGGTTTGCACTGCCGAGGTTTGCACTGCC GAGGTTTGCACTGCCGAGGTTTGCACTGCC	TTGCCGAGGTTTGCACTGCCGAGGTTTGCACTGCC GAGGTTTGCACTGCC	Codon deletion	ggcagtgcaaacctc/-	PPE8	GSANL2436-
	510831	TCCGGGGGGCGCACCGGGGG	TCCGGGGG	Codon deletion	tcccccggtgcgccc/tcc	ctpH	SPGAP1491S
	580773	GGGGGCACCACCCGCTTGCGGGGGAG	GG	Intergenic			
	769962	GCCC	GCC	Frameshift		end	
	859131	C	CA	Frameshift		cyp123	
	874835	CC	CCGC	Frameshift		ptrBa	
	889014	AAGGAGG	A	Intergenic			
	968426	AGCCGGGTTGGCCGG	AGCCGGGTTGGCCGGGTTGGCCGG	Codon insertion	-/CCGGCCAAC	PE_PGRS15	-602PAN
	976896	TTG	T	Frameshift		PPE13	
	1010204	CGG	CGGG	Frameshift		Rv0907	
	1150321	TGCCCGACAGCAAGGCC	TGCC	Frameshift		kdpD	
	1159161	GCCCACC	GCC	Frameshift		Rv1034c	
	1165521	TA	TAA	Intergenic			
	1168715	CTT	CTTT	Frameshift		Rv1046c	
	1277869	GT	GTT	Intergenic			
	1313337	A	AG	Intergenic			
	1365837	CGGGG	CGGGGG	Intergenic			
	1374936	CAA	CA	Frameshift		Rv1232c	
	1527449	GT	GTT	Frameshift		Rv1358	
	1625347	CGGG	CGGGG	Frameshift		opcA	
	1753519	GCCCC	GCCCCC	Frameshift		fadD11.1	
	1894300	GG	GGTCTTGCCGCG	Frameshift		Rv1668c	
	1986623	GTT	GTTT	Frameshift		Rv1754c	
	2094911	ACAGCGTCAGCGTCAG	ACAGCGTCAG	Codon deletion	actgacgct/act	gnd1	TDA88T
	2109523	CGG	CGGG	Intergenic			
	2133468	TTCGCAT	TTCGCATGCCGTCACCTCGCAT	Nonsense	gta/gATGCGAGGTGACGGCta	Rv1883c	V73DAR*RL
	2137521	AC	ACTCCGATCACC	Frameshift		Rv1888c	
	2207591	T	TC	Intergenic			
	2268725	A	AG	Frameshift		Rv2024c	
	2300205	A	AC	Frameshift		pks12	

2357268 TGCCGCCGCCGCCGCCGCCG	TGCCGCCGCCGCCGCCG	Frameshift		PE_PGRS36	
2365413 AT	ATT	Intergenic			
2368564 TAA	TA	Intergenic			
2382085 AGTG	AG	Frameshift		PPE37	
2401883 C	T	Intergenic			
2470886 AAACTTAGCCGCGA	AA	Codon change and Codon deletion	aaacttagccgcgag/aag	Rv2206	KLSRE89K
2523205 GCG	GCGCCG	Intergenic			
2525722 CGGGGG	CGGGG	Frameshift		Rv2250A	
2617632 A	AT	Intergenic			
2704884 AC	ACAGCGACCATATCGCCGAGCTC	Codon insertion	-/AGCGACCATATCGCCGAGCTC	Rv2407	-64SDHIAEL
2881597 AGGGGGG	AGGGGG	Frameshift		Rv2561	
3028101 ACC	ACCC	Frameshift		Rv2715	
3073868 T	C	Missense	Acc/Gcc	thyA	T202A
3131469 TTGT	TTGTCGGCGATGT	Codon insertion	-/ACATCGCCG	Rv2823c	-101TSP
3179242 ACGGCGGCGGCG	ACGGCGGCG	Codon deletion	gacgcc/gac	Rv2867c	DA20D
3190145 TCCCCCC	TCCCCC	Frameshift		Rv2880c	
3247856 G	C	Synonymous	cgG/cgC	ppsA	R804
3302324 AGCGAG	A	Frameshift		fadD29	
3415180 ACACCTAGGGGGTGGCACCTAGGGGGTGGCAC	ACACCTAGGGGGTGGCAC	Intergenic			
3473996 GA	GAA	Intergenic			
3480375 GAAA	GAA	Frameshift		Rv3113	
3519294 GGCC	GGCCGCC	Codon change and Codon insertion	cac/cGCCac	nuoH	H6RH
3554066 C	CG	Intergenic			
3580636 CTTT	CTT	Frameshift		lipV	
3590686 GT	GCT	Intergenic			
3595488 CAGCGTAGT	CAGCGTAGTCTGGGAGCGTAGT	Intergenic			
3730741 G	A	Synonymous	ggC/ggT	PPE54	G2065
3884906 A	G	Missense	Aaa/Gaa	Rv3467	K315E
3998059 AGGCGGCGGCGGC	AGGCGGCGGC	Codon change and Codon deletion	gcggcc/gcc	PPE64	AA30A
4093879 TGGGGG	TGGGG	Frameshift		PE_PGRS60	
4095001 CG	C	Frameshift		Rv3655c	
4197138 CTT	CTTT	Intergenic			
4198611 CGGGGG	CGGGG	Intergenic			
4336090 A	AT	Intergenic			
4338595 GC	G	Intergenic			
4353201 CCCGCCGCCGCC	CCCGCCGCC	Codon change and Codon deletion	ccgcca/cca	Rv3876	PP67P
4383144 CC	CCGGGGC	Frameshift		Rv3897c	
4400660 ACCC	ACC	Frameshift		sigM	

[a] Mutated nucleotide is indicated by capitalized letter.

Table S7.4 - Mutations found to be acquired along node-delimited branches in the Q1 subtree. Position, Reference Sequence and Mutated Sequence are derived from the VCF format.

Branch	Position	Reference Sequence	Mutated Sequence	Type	Codon Change[a]	Affected ORF	Mutation
A-B							
	49439	C	T	Synonymous	gcG/gcA	Rv0045c	A167
	176432	A	G	Missense	Acc/Gcc	Rv0149	T245A
	402528	C	T	Missense	tGc/tAc	aspC	C212Y
	501882	C	G	Synonymous	gtC/gtG	thiO	V245
	661798	A	G	Synonymous	ttA/ttG	nrdZ	L168
	912267	C	G	Missense	Ccg/Gcg	Rv0819	P178A
	1187439	T	C	Missense	gTt/gCt	Rv1065	V2A
	1267195	G	A	Intergenic			
	1267253	G	A	Intergenic			
	1295129	C	T	Missense	aCc/aTc	typA	T321I
	1673425	C	T	Intergenic			
	1771865	C	T	Synonymous	cgG/cgA	Rv1565c	R655
	1790063	C	T	Intergenic			
	1877638	A	G	Missense	Act/Gct	pks7	T779A
	1973157	T	G	Synonymous	ccT/ccG	pknF	P340
	2058302	A	G	Missense	cAg/cGg	Rv1816	Q16R
	2088604	G	A	Synonymous	gaC/gaT	PE_PGRS34	D305
	2204092	G	A	Missense	aCc/aTc	Rv1960c	T45I
	2211971	G	C	Missense	Gac/Cac	mce3C	D116H
	2300639	G	A	Synonymous	acC/acT	pks12	T2116
	2439519	G	A	Synonymous	cgC/cgT	Rv2177c	R143
	2584866	G	C	Intergenic			
	2684464	G	A	Intergenic			
	2701886	T	G	Missense	Acc/Ccc	lepA	T455P
	2780470	A	C	Missense	cTg/cGg	gdh	L598R
	2964540	G	A	Missense	Ggc/Agc	Rv2638	G46S
	2974933	A	G	Missense	aTc/aCc	Rv2650c	I101T
	3401605	G	A	Missense	gCg/gTg	Rv3041c	A105V
	3593546	T	C	Synonymous	aaT/aaC	Rv3216	N9
	3694960	C	T	Missense	Cgc/Tgc	pmmB	R33C
	3792581	G	C	Missense	aCg/aGg	Rv3378c	T223R
	3992801	C	G	Synonymous	gtC/gtG	fdxB	V39
	4097203	G	C	Missense	aaC/aaG	Rv3659c	N264K
	4266215	C	T	Synonymous	ctG/ctA	fbpA	L148
	4277856	C	T	Missense	cCg/cTg	PE_PGRS62	P429L
	4365983	T	C	Missense	Acc/Gcc	Rv3884c	T286A
B-C							
	15505	G	C	Missense	Gac/Cac	trpG	D198H
	105060	G	A	Synonymous	gaC/gaT	Rv0095c	D52
	164643	T	C	Synonymous	gtT/gtC	cyp138	V426
	204604	C	T	Synonymous	atC/atT	lprK	I180
	236250	G	A	Synonymous	agC/agT	Rv0198c	S86
	490655	T	C	Synonymous	agA/agG	Rv0406c	R17
	507624	T	C	Synonymous	cgA/cgG	Rv0421c	R46
	666566	A	C	Missense	aTg/aGg	Rv0573c	M226R
	1156979	C	A	Missense	cGg/cTg	trcS	R326L
	1274156	C	T	Synonymous	Ctg/Ttg	mmpL13b	L268
	1324915	T	G	Missense	Aat/Cat	Rv1184c	N233H
	1465786	C	T	Synonymous	ctC/ctT	atpG	L301
	1598720	C	A	Missense	gCg/gAg	whiA	A272E
	1674782	T	C	Missense	aTc/aCc	inhA	I194T
	1777422	C	T	Synonymous	Ctg/Ttg	bioF1	L241
	1862460	T	C	Missense	Ata/Gta	PE_PGRS30	I975V
	1888061	A	G	Synonymous	gcA/gcG	pks9	A12
	1923027	C	T	Missense	Cac/Tac	Rv1698	H95Y
	2009774	CGGG	CGGGG	Frameshift		Rv1775	
	2009908	A	G	Missense	aAt/aGt	Rv1775	N246S
	2041559	GC	G	Intergenic			
	2193376	C	G	Missense	gaC/gaG	ribA1	D257E
	2443423	C	T	Missense	aCg/aTg	Rv2181	T41M
	2854654	G	A	Synonymous	atC/atT	Rv2530c	I11
	3007095	A	G	Intergenic			
	3007148	A	G	Intergenic			
	3028101	ACC	ACCC	Frameshift		Rv2715	
	3138416	G	A	Missense	Ccg/Tcg	ugpC	P256S
	3141353	C	G	Synonymous	acG/acC	ugpA	T290
	3520172	G	A	Synonymous	ctG/ctA	nuoH	L297
	3652033	TG	TGG	Frameshift		ctpC	
	3680075	C	T	Missense	Ccg/Tcg	lhr	P1101S
	4031470	G	T	Synonymous	gtC/gtA	PE_PGRS58	V563
	4182269	C	G	Missense	tCc/tGc	ligC	S171C
	4186821	T	C	Missense	aTg/aCg	Rv3736	M63T
	4195009	G	A	Missense	gCa/gTa	ctpJ	A122V
	4198319	A	G	Synonymous	ctT/ctC	Rv3750c	L93
	4216505	C	A	Missense	Gcg/Tcg	Rv3771c	A76S
	4225830	G	A	Synonymous	cgG/cgA	Rv3779	R282
	4247781	T	C	Missense	aTg/aCg	embB	M423T
	4328127	G	C	Synonymous	tcG/tcC	ethR	S193
	4402703	A	C	Missense	Agt/Cgt	trxB2	S326R
C-D							
	484379	G	T	Missense	Gcg/Tcg	fadD30	A135S
	761155	C	T	Missense	tCg/tTg	rpoB	S450L
	2288868	A	C	Missense	gTc/gGc	pncA	V125G
	3550815	A	C	Missense	gAt/gCt	Rv3183	D34A
	3862472	GAA	GA	Intergenic			
	4247429	A	G	Missense	Atg/Gtg	embB	M306V
D-E							
	761998	T	C	Missense	cTg/cCg	rpoB	L731P
E-F							
	1473246	A	G	rRNA mutation	-	rrs	
F-G							
	836538	A	G	Missense	Aac/Gac	PE_PGRS9	N280D
		AGCCGGGTTGGCCGG	AGCCGGGTTGGCCGGGTTGGCCG				
	968426		G	Codon insertion	-/CCGGCCAAC	PE_PGRS15	-602PAN
	2365413	AT	ATT	Intergenic			

[a] Mutated nucleotide is indicated by capitalized letter.

Table S7.5 - Types and number of large SVs (≥100 bp) found among the 75 analyzed isolates using the SVMerge pipeline and local assembly validation.

SV Type	No. of variants	Size range (bp)
Deletion	552	100-26351
Deletion + Insertion	8	540-1714
Insertion (incompletely reconstructed)	42	138-1822
Insertion (completely reconstructed)	109	106-562
Inversion	32	258-10493
Complex Inversion	12	1308-8088
Invertion + Deletion	19	1224-9617
Inversion + Insertion	15	212-7114

Table S7.6 - Multiple comparison test results upon comparison of mean overall N_s/S and T_v/T_s ratios for four groups of strains: Lisboa3, Q1, Beijing clades and, other non clustered strains (NC). Significant differences at the 0.05 level are highlighted in bold.

Clades		N_s/S Ratio		T_v/T_s Ratio	
A	**B**	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***
Lisboa3	**Q1**	0.0249	0.4112	-0.0455	**0.0330**
	Beijing	-0.1156	**0.0146**	-0.0642	0.1066
	NC	-0.0792	**0.0026**	-0.0105	0.9773
Q1	**Lisboa3**	-0.0249	0.4112	0.0455	**0.0330**
	Beijing	-0.1405	**0.0038**	-0.0188	0.9673
	NC	-0.1042	**0.0001**	0.0350	0.2256
Beijing	**Lisboa3**	0.1156	**0.0146**	0.0642	0.1066
	Q1	0.1405	**0.0038**	0.0188	0.9673
	NC	0.0364	0.8487	0.0537	0.2410
NC	**Lisboa3**	0.0792	**0.0026**	0.0105	0.9773
	Q1	0.1042	**0.0001**	-0.0350	0.2256
	Beijing	-0.0364	0.8487	-0.0537	0.2410

Table S7.7 - Multiple comparison test results upon comparison of mean N_S/S and T_V/T_S ratios across the four genomic quadrants. Significant differences at the 0.05 level are highlighted in bold.

Quadrant		N_S/S Ratio		T_V/T_S Ratio	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
1	2	.3933	**.0000**	-.0982	**.0000**
	3	.1041	**.0001**	.0366	.0987
	4	.0856	**.0083**	-.0044	.9995
2	1	-.3933	**.0000**	.0982	**.0000**
	3	-.2892	**.0000**	.1348	**.0000**
	4	-.3078	**.0000**	.0938	**.0000**
3	1	-.1041	**.0001**	-.0366	.0987
	2	.2892	**.0000**	-.1348	**.0000**
	4	-.0185	.9871	-.0411	.0614
4	1	-.0856	**.0083**	.0044	.9995
	2	.3078	**.0000**	-.0938	**.0000**
	3	.0185	.9871	.0411	.0614

Table S7.8 - Multiple comparison test results upon comparison of mean Ns/S and Tv/Ts ratios across the four genomic quadrants for four groups of strains: Lisboa3, Q1, Beijing clades and, other non clustered strains (NC). Significant differences at the 0.05 level are highlighted in bold.

Clades		Quadrant															
		Quadrant 1				Quadrant 2				Quadrant 3				Quadrant 4			
		N_s/S Ratio		T_v/T_s Ratio		N_s/S Ratio		T_v/T_s Ratio		N_s/S Ratio		T_v/T_s Ratio		N_s/S Ratio		T_v/T_s Ratio	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	-0.0850	0.3236	-0.0497	0.3818	-0.1302	**0.0000**	0.0409	1.0000	0.0348	0.8037	-0.1787	**0.0000**	0.3244	**0.0000**	-0.1787	**0.0000**
	Beijing	0.0673	1.0000	-0.0509	0.5228	-0.3825	**0.0000**	-0.0726	0.2574	-0.1224	0.2409	-0.1804	**0.0001**	0.0326	0.9518	-0.1804	**0.0001**
	NC	-0.0572	0.6335	-0.0158	1.0000	-0.2337	**0.0000**	0.0184	1.0000	0.0011	1.0000	-0.0872	**0.0180**	0.0224	0.9939	-0.0872	**0.0180**
Q1	Lisboa3	0.0850	0.3236	0.0497	0.3818	0.1302	**0.0000**	-0.0409	1.0000	-0.0348	0.8037	0.1787	**0.0000**	-0.3244	**0.0000**	0.1787	**0.0000**
	Beijing	0.1523	**0.0183**	-0.0012	1.0000	-0.2522	**0.0000**	-0.1135	**0.0150**	-0.1572	0.1014	-0.0017	1.0000	-0.2918	**0.0000**	-0.0017	1.0000
	NC	0.0278	1.0000	0.0339	0.8065	-0.1035	**0.0068**	-0.0225	1.0000	-0.0337	0.9465	0.0915	**0.0186**	-0.3020	**0.0000**	0.0915	**0.0186**
Beijing	Lisboa3	-0.0673	1.0000	0.0509	0.5228	0.3825	**0.0000**	0.0726	0.2574	0.1224	0.2409	0.1804	**0.0001**	-0.0326	0.9518	0.1804	**0.0001**
	Q1	-0.1523	**0.0183**	0.0012	1.0000	0.2522	**0.0000**	0.1135	**0.0150**	0.1572	0.1014	0.0017	1.0000	0.2918	**0.0000**	0.0017	1.0000
	NC	-0.1245	**0.0271**	0.0351	1.0000	0.1488	**0.0038**	0.0910	**0.0266**	0.1234	0.2940	0.0932	0.0514	-0.0102	1.0000	0.0932	0.0514
NC	Lisboa3	0.0572	0.6335	0.0158	1.0000	0.2337	**0.0000**	-0.0184	1.0000	-0.0011	1.0000	0.0872	**0.0180**	-0.0224	0.9939	0.0872	**0.0180**
	Q1	-0.0278	1.0000	-0.0339	0.8065	0.1035	**0.0068**	0.0225	1.0000	0.0337	0.9465	-0.0915	**0.0186**	0.3020	**0.0000**	-0.0915	**0.0186**
	Beijing	0.1245	**0.0271**	-0.0351	1.0000	-0.1488	**0.0038**	-0.0910	**0.0266**	-0.1234	0.2940	-0.0932	0.0514	0.0102	1.0000	-0.0932	0.0514

Table S7.9 - Multiple comparison test results upon comparison of mean N_s/S ratios across the different COGs for four groups of strains: Lisboa3, Q1, Beijing clades and, other non clustered strains (NC). Significant differences at the 0.05 level are highlighted in bold.

Clades		COG[a]													
		C		D		E		F		G		H		I	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	0.1074	**0.0000**	0.2821	**0.0292**	-0.0869	**0.0029**	0.0631	**0.0448**	0.0635	0.9491	-0.1717	0.5064	-0.1300	**0.0088**
	NC	-0.1415	**0.0026**	0.8678	**0.0000**	-0.2935	**0.0001**	-0.1314	**0.0018**	-0.6906	0.2013	-0.2673	0.2588	-0.2348	**0.0012**
	Beijing	-0.2203	**0.0002**	0.7458	**0.0088**	-0.1664	0.2128	-0.4434	**0.0000**	-0.7993	**0.0338**	0.4273	**0.0008**	-0.3617	**0.0000**
Q1	Lisboa3	-0.1074	**0.0000**	-0.2821	**0.0292**	0.0869	**0.0029**	-0.0631	**0.0448**	-0.0635	0.9491	0.1717	0.5064	0.1300	**0.0088**
	NC	-0.2489	**0.0000**	0.5858	**0.0001**	-0.2066	**0.0081**	-0.1944	**0.0000**	-0.7540	0.1505	-0.0956	0.9844	-0.1048	0.2097
	Beijing	-0.3278	**0.0000**	0.4638	0.1068	-0.0796	0.8063	-0.5065	**0.0000**	-0.8628	**0.0213**	0.5990	**0.0001**	-0.2317	**0.0025**
NC	Lisboa3	0.1415	**0.0026**	-0.8678	**0.0000**	0.2935	**0.0001**	0.1314	**0.0018**	0.6906	0.2013	0.2673	0.2588	0.2348	**0.0012**
	Q1	0.2489	**0.0000**	-0.5858	**0.0001**	0.2066	**0.0081**	0.1944	**0.0000**	0.7540	0.1505	0.0956	0.9844	0.1048	0.2097
	Beijing	-0.0789	0.4273	-0.1220	0.9762	0.1270	0.6569	-0.3120	**0.0001**	-0.1088	0.9999	0.6946	**0.0000**	-0.1268	0.2564
Beijing	Lisboa3	0.2203	**0.0002**	-0.7458	**0.0088**	0.1664	0.2128	0.4434	**0.0000**	0.7993	**0.0338**	-0.4273	**0.0008**	0.3617	**0.0000**
	Q1	0.3278	**0.0000**	-0.4638	0.1068	0.0796	0.8063	0.5065	**0.0000**	0.8628	**0.0213**	-0.5990	**0.0001**	0.2317	**0.0025**
	NC	0.0789	0.4273	0.1220	0.9762	-0.1270	0.6569	0.3120	**0.0001**	0.1088	0.9999	-0.6946	**0.0000**	0.1268	0.2564

[a]no mutations were found in COG A genes.

Clades		COG[a]													
		J		K		L		M		N		O		P	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	0.1190	0.0659	0.0970	**0.0255**	-0.5724	0.6045	-0.0301	1.0000	0.0834	0.2218	-0.3274	**0.0000**	0.4746	**0.0000**
	NC	0.0048	1.0000	-0.1260	0.5482	-0.6868	0.0910	0.3619	**0.0342**	-0.3275	**0.0000**	-0.3546	**0.0000**	0.5538	**0.0000**
	Beijing	0.0468	0.9647	-0.4959	**0.0039**	0.2742	1.0000	1.0768	**0.0000**	-0.7615	**0.0001**	-1.1524	**0.0001**	1.0010	**0.0000**
Q1	Lisboa3	-0.1190	0.0659	-0.0970	**0.0255**	0.5724	0.6045	0.0301	1.0000	-0.0834	0.2218	0.3274	**0.0000**	-0.4746	**0.0000**
	NC	-0.1142	0.4149	-0.2230	**0.0479**	-0.1144	1.0000	0.3919	0.1776	-0.4109	**0.0000**	-0.0272	0.9980	0.0791	0.9374
	Beijing	-0.0721	0.7210	-0.5929	**0.0013**	0.8466	0.2502	1.1068	**0.0000**	-0.8449	**0.0000**	-0.8250	**0.0005**	0.5264	**0.0000**
NC	Lisboa3	-0.0048	1.0000	0.1260	0.5482	0.6868	0.0910	-0.3619	**0.0342**	0.3275	**0.0000**	0.3546	**0.0000**	-0.5538	**0.0000**
	Q1	0.1142	0.4149	0.2230	**0.0479**	0.1144	1.0000	-0.3919	0.1776	0.4109	**0.0000**	0.0272	0.9980	-0.0791	0.9374
	Beijing	0.0420	0.9947	-0.3699	**0.0331**	0.9610	**0.0484**	0.7149	**0.0000**	-0.4339	**0.0014**	-0.7978	**0.0002**	0.4473	**0.0004**
Beijing	Lisboa3	-0.0468	0.9647	0.4959	**0.0039**	-0.2742	1.0000	-1.0768	**0.0000**	0.7615	**0.0001**	1.1524	**0.0001**	-1.0010	**0.0000**
	Q1	0.0721	0.7210	0.5929	**0.0013**	-0.8466	0.2502	-1.1068	**0.0000**	0.8449	**0.0000**	0.8250	**0.0005**	-0.5264	**0.0000**
	NC	-0.0420	0.9947	0.3699	**0.0331**	-0.9610	**0.0484**	-0.7149	**0.0000**	0.4339	**0.0014**	0.7978	**0.0002**	-0.4473	**0.0004**

Table S7.9 (cont)

Clades		COG[a]													
		Q		R		S		T		U		V		Not in COG	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	0.0562	0.4277	0.5745	**0.0000**	0.0544	0.8148	-0.2211	1.0000	0.0000	-	-2.7872	**0.0000**	-0.0508	0.3852
	NC	-0.0178	0.9965	0.1572	0.1308	-0.1377	**0.0306**	-0.4549	1.0000	0.1789	0.9296	-0.4342	0.3460	-0.0784	0.0740
	Beijing	-0.2713	**0.0001**	0.3293	**0.0001**	-0.1053	0.7991	-0.9288	0.3723	-0.5714	0.3363	-0.4875	0.3604	-0.0139	1.0000
Q1	Lisboa3	-0.0562	0.4277	-0.5745	**0.0000**	-0.0544	0.8148	0.2211	1.0000	0.0000	-	2.7872	**0.0000**	0.0508	0.3852
	NC	-0.0740	0.2977	-0.4174	**0.0000**	-0.1921	**0.0100**	-0.2338	1.0000	0.1789	0.9296	2.3529	**0.0000**	-0.0276	0.9632
	Beijing	-0.3275	**0.0000**	-0.2452	**0.0018**	-0.1597	0.4925	-0.7077	0.9821	-0.5714	0.3363	2.2997	**0.0000**	0.0369	0.9928
NC	Lisboa3	0.0178	0.9965	-0.1572	0.1308	0.1377	**0.0306**	0.4549	1.0000	-0.1789	0.9296	0.4342	0.3460	0.0784	0.0740
	Q1	0.0740	0.2977	0.4174	**0.0000**	0.1921	**0.0100**	0.2338	1.0000	-0.1789	0.9296	-2.3529	**0.0000**	0.0276	0.9632
	Beijing	-0.2535	**0.0001**	0.1722	0.1000	0.0324	0.9994	-0.4739	1.0000	-0.7504	0.2214	-0.0533	1.0000	0.0644	0.9130
Beijing	Lisboa3	0.2713	**0.0001**	-0.3293	**0.0001**	0.1053	0.7991	0.9288	0.3723	0.5714	0.3363	0.4875	0.3604	0.0139	1.0000
	Q1	0.3275	**0.0000**	0.2452	**0.0018**	0.1597	0.4925	0.7077	0.9821	0.5714	0.3363	-2.2997	**0.0000**	-0.0369	0.9928
	NC	0.2535	**0.0001**	-0.1722	0.1000	-0.0324	0.9994	0.4739	1.0000	0.7504	0.2214	0.0533	1.0000	-0.0644	0.9130

Table S7.10 - Multiple comparison test results upon comparison of mean Tv/Ts ratios across the different COGs for four groups of strains: Lisboa3, Q1, Beijing clades and, other non clustered strains (NC). Significant differences at the 0.05 level are highlighted in bold.

Clades		COG[a]													
		C		D		E		F		G		H		I	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	0.0073	1.0000	0.1718	0.4818	-0.0890	**0.0061**	0.0168	0.9100	0.0315	0.8753	-0.2620	**0.0000**	-0.0877	0.2533
	NC	-0.0426	0.3491	-0.0107	1.0000	-0.0555	0.2272	-0.1100	**0.0097**	-0.0181	0.9977	-0.1940	**0.0000**	0.0085	1.0000
	Beijing	-0.2082	**0.0005**	0.3312	**0.0410**	-0.1057	0.2348	-0.1346	**0.0089**	0.0673	0.3255	-0.2377	**0.0002**	-0.1501	**0.0186**
Q1	Lisboa3	-0.0073	1.0000	-0.1718	0.4818	0.0890	**0.0061**	-0.0168	0.9100	-0.0315	0.8753	0.2620	**0.0000**	0.0877	0.2533
	NC	-0.0498	0.7286	-0.1825	0.4990	0.0335	0.8444	-0.1269	**0.0015**	-0.0496	0.4735	0.0680	0.4895	0.0963	0.0540
	Beijing	-0.2155	**0.0009**	0.1594	0.7889	-0.0166	0.9994	-0.1515	**0.0043**	0.0358	0.6304	0.0244	0.9861	-0.0623	1.0000
NC	Lisboa3	0.0426	0.3491	0.0107	1.0000	0.0555	0.2272	0.1100	**0.0097**	0.0181	0.9977	0.1940	**0.0000**	-0.0085	1.0000
	Q1	0.0498	0.7286	0.1825	0.4990	-0.0335	0.8444	0.1269	**0.0015**	0.0496	0.4735	-0.0680	0.4895	-0.0963	0.0540
	Beijing	-0.1656	**0.0025**	0.3419	**0.0385**	-0.0501	0.8882	-0.0246	0.9907	0.0854	0.1058	-0.0436	0.8900	-0.1586	**0.0030**
Beijing	Lisboa3	0.2082	**0.0005**	-0.3312	**0.0410**	0.1057	0.2348	0.1346	**0.0089**	-0.0673	0.3255	0.2377	**0.0002**	0.1501	**0.0186**
	Q1	0.2155	**0.0009**	-0.1594	0.7889	0.0166	0.9994	0.1515	**0.0043**	-0.0358	0.6304	-0.0244	0.9861	0.0623	1.0000
	NC	0.1656	**0.0025**	-0.3419	**0.0385**	0.0501	0.8882	0.0246	0.9907	-0.0854	0.1058	0.0436	0.8900	0.1586	**0.0030**

[a]no mutations were found in COG A genes; not possible to perform for COG U since some categories had fewer than two cases.

Clades		J		K		L		M		N		O		P	
A	B	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*	Mean Difference (A-B)	*p*
Lisboa3	Q1	-0.4011	**0.0410**	-0.1486	0.4345	-1.1562	**0.0001**	-0.2954	**0.0000**	0.0264	1.0000	-0.0250	0.9864	-0.1568	**0.0000**
	NC	0.0256	0.9999	-0.0617	1.0000	-0.1015	0.9348	0.0395	0.6106	0.1588	**0.0137**	0.0748	0.2099	-0.2175	**0.0001**
	Beijing	0.1553	0.4084	-0.1953	0.2497	0.3457	**0.0014**	-0.0006	1.0000	0.4371	**0.0000**	0.0846	0.1638	-0.0086	1.0000
Q1	Lisboa3	0.4011	**0.0410**	0.1486	0.4345	1.1562	**0.0001**	0.2954	**0.0000**	-0.0264	1.0000	0.0250	0.9864	0.1568	**0.0000**
	NC	0.4267	**0.0113**	0.0869	1.0000	1.0546	**0.0001**	0.3349	**0.0000**	0.1324	0.0875	0.0998	0.1388	-0.0606	0.6949
	Beijing	0.5565	**0.0012**	-0.0467	1.0000	1.5018	**0.0000**	0.2948	**0.0000**	0.4107	**0.0000**	0.1096	0.1059	0.1482	**0.0219**
NC	Lisboa3	-0.0256	0.9999	0.0617	1.0000	0.1015	0.9348	-0.0395	0.6106	-0.1588	**0.0137**	-0.0748	0.2099	0.2175	**0.0001**
	Q1	-0.4267	**0.0113**	-0.0869	1.0000	-1.0546	**0.0001**	-0.3349	**0.0000**	-0.1324	0.0875	-0.0998	0.1388	0.0606	0.6949
	Beijing	0.1297	0.0754	-0.1337	0.6823	0.4472	**0.0002**	-0.0401	0.8351	0.2783	**0.0003**	0.0098	0.9999	0.2088	**0.0046**
Beijing	Lisboa3	-0.1553	0.4084	0.1953	0.2497	-0.3457	**0.0014**	0.0006	1.0000	-0.4371	**0.0000**	-0.0846	0.1638	0.0086	1.0000
	Q1	-0.5565	**0.0012**	0.0467	1.0000	-1.5018	**0.0000**	-0.2948	**0.0000**	-0.4107	**0.0000**	-0.1096	0.1059	-0.1482	**0.0219**
	NC	-0.1297	0.0754	0.1337	0.6823	-0.4472	**0.0002**	0.0401	0.8351	-0.2783	**0.0003**	-0.0098	0.9999	-0.2088	**0.0046**

Table 7.10 (cont)

Clades		Q		R		S		T		V		Not in COG	
A	**B**	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***	**Mean Difference (A-B)**	***p***
Lisboa3	**Q1**	-0.0904	**0.0017**	-0.0328	0.5666	-0.1114	0.2938	-0.1139	0.1037	-0.3456	**0.0001**	0.0595	**0.0011**
	NC	-0.0589	0.2656	0.0571	**0.0339**	0.0398	1.0000	0.1728	**0.0068**	-0.5103	**0.0001**	-0.0171	0.9612
	Beijing	-0.0924	0.4650	-0.1261	**0.0123**	-0.1703	0.0594	0.2943	**0.0428**	-0.4182	**0.0001**	-0.1598	0.1011
Q1	**Lisboa3**	0.0904	**0.0017**	0.0328	0.5666	0.1114	0.2938	0.1139	0.1037	0.3456	**0.0001**	-0.0595	**0.0011**
	NC	0.0315	0.8461	0.0899	**0.0148**	0.1512	**0.0117**	0.2867	**0.0002**	-0.1646	0.6417	-0.0766	**0.0034**
	Beijing	-0.0020	1.0000	-0.0932	0.0882	-0.0589	1.0000	0.4082	**0.0060**	-0.0725	0.8580	-0.2194	**0.0235**
NC	**Lisboa3**	0.0589	0.2656	-0.0571	**0.0339**	-0.0398	1.0000	-0.1728	**0.0068**	0.5103	**0.0001**	0.0171	0.9612
	Q1	-0.0315	0.8461	-0.0899	**0.0148**	-0.1512	**0.0117**	-0.2867	**0.0002**	0.1646	0.6417	0.0766	**0.0034**
	Beijing	-0.0335	0.9888	-0.1831	**0.0004**	-0.2101	**0.0026**	0.1215	0.7305	0.0921	0.9591	-0.1427	0.1646
Beijing	**Lisboa3**	0.0924	0.4650	0.1261	**0.0123**	0.1703	0.0594	-0.2943	**0.0428**	0.4182	**0.0001**	0.1598	0.1011
	Q1	0.0020	1.0000	0.0932	0.0882	0.0589	1.0000	-0.4082	**0.0060**	0.0725	0.8580	0.2194	**0.0235**
	NC	0.0335	0.9888	0.1831	**0.0004**	0.2101	**0.0026**	-0.1215	0.7305	-0.0921	0.9591	0.1427	0.1646

Table S7.11 – End sequences from the different ISs used as probes to extract reads for mapping analysis.

IS	5' End	3' End
IS6110	TGAACCGCCCCGGCATGTCCGGAGA	TCTCCGGACTCACCGGGGCGGTTCA
IS1532	GTGGCCACGATAGCCCAACGGCTGC	CAAGAGACCCGGCCGCACCACCAGC
IS1533	AATGTTGACTGTGGAAGATTGGGCT	CCACCAGCCGGAACCACCGAAGAAT
IS1537	GTTTTGAGTGGTCGGTAGTTGTCGG	ACTAAAGATCGCTCGCTTGCAACGG
IS1538	TGAGTGGTCGGTAGTTGTCAGTGGCGTGCG	none[a]
IS1539	ATGTCAAGGATACTGACACATGTGC	CCGAGACGGGGTGCAAGTCAAGTGA
IS1540	CGTGATCGACACGGCCATCGAGGAG	CTCCTGCCATCGGGGCAGATACGCT
IS1552	GGTGTTTGCCGAGCTGATCCGTGCC	GTCGCCGACCCTCTCTGTCAACCTT
IS1553	GAGATCGTCGGTGTGTACCGCCAAG	GCGGCCAGTAGGCACCGACGAACTC
IS1556	TGACACATGACCGCTTTCCGCCGCC	TTTTGCGGTTGATGGATGGTAGCTG
IS1560	TTGATCCCCGGTCGGATGGTGCTGA	GGTCACGATCAGCGCCCTCCCCGCA
IS1561	GTGGCCATCGACCCCGCCGCTGCCT	none[a]
IS1602	none[a]	AGCACGCCACTGACAACTACCGACCACTCA
IS1603	GTGCGCCAAATTAGTAGTCGCTATC	ACCAGACCATCAATTGTTGCGACGT
IS1604	GGTGGCGGTCGGCGATGACGAGGAG	ACCGGCCAGGAGGCCCAACCGAAAT
IS1605	TTGGGCCCATCGTCGAAAACCTGCC	TAGTCGCGTGACCACTAAAGATCACTCACTTGCAACGG
IS1606	GGTGACCCGCGATCCACACAGCCCG	ATCGAGCCTTAAGGAAAATCAAAGC
IS1607	GCGGCCCAAACCACTACCCTGCCCG	CCTCCAGGCACTGCTCAACCAAGAT
IS1608	GTGACTGCCGAGAATCCAGGACGGA	GCGTCGGTTGATGTCGTTGGCGCGG
ISB9	ATCACCCCGGCAAGCTGTCCCTGAG	GCCGAGCATCTCTTTGCGGGGTGAT

[a] Sequences at this end did not allow the unambiguous mapping of the respective Insertion Sequence